

# PATENT ABSTRACTS OF JAPAN

(11) Publication number : 2003-101939

(43) Date of publication of application : 04.04.2003

(51) Int.Cl.

H04N 5/91  
G06F 17/30  
G10L 11/02  
G10L 15/00  
G10L 15/04  
H04N 5/76

(21) Application number : 2001-304360

(71) Applicant : PIONEER ELECTRONIC CORP

(22) Date of filing : 28.09.2001

(72) Inventor : NAKAMURA TAKESHI

(30) Priority

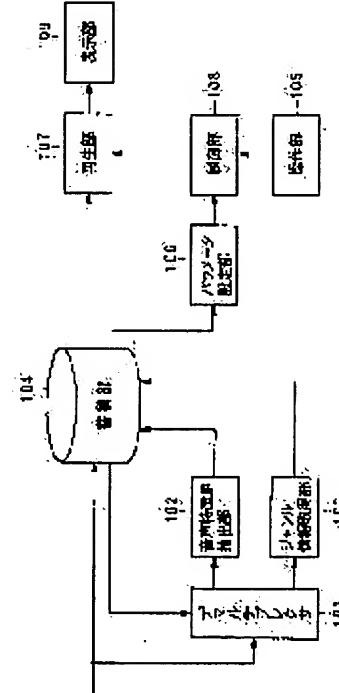
Priority number : 2001216604 Priority date : 17.07.2001 Priority country : JP

## (54) APPARATUS, METHOD, AND PROGRAM FOR SUMMARIZING VIDEO INFORMATION

(57) Abstract:

**PROBLEM TO BE SOLVED:** To provide a summary reproducing device capable of accurately reproducing a summary for each type of video information and of reducing a burden of generating processing of digest information.

**SOLUTION:** This summary reproducing apparatus 100 comprises a sound feature amount extracting part 102 for obtaining a sound feature amount on the basis of a preset parameter from entered audio/video information, a genre information obtaining part 103 for obtaining genre information from additional information added to the entered audio/video information, a decision parameter setting part 106 for setting an optimum parameter for extracting a sound feature amount on the basis of genre information, and a controlling part 108 for deciding digest segments to be extracted in stored audio/video information on the basis of a sound feature amount suitable for the preset parameter and for controlling a reproducing part 197 on the basis of the digest segments, wherein a summary is reproduced by using a parameter optimized for each genre.



---

**LEGAL STATUS**

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-101939

(P2003-101939A)

(43)公開日 平成15年4月4日(2003.4.4)

(51)Int.Cl.  
H 04 N 5/91  
G 06 F 17/30  
G 10 L 11/02

識別記号  
170  
220

F I  
G 06 F 17/30  
H 04 N 5/76  
5/91

コード(参考)  
170 D 5 B 075  
170 E 5 C 052  
220 A 5 C 053  
B 5 D 015  
N

審査請求 未請求 請求項の数30 OL (全32頁) 最終頁に続く

(21)出願番号 特願2001-304360(P2001-304360)  
(22)出願日 平成13年9月28日(2001.9.28)  
(31)優先権主張番号 特願2001-216604(P2001-216604)  
(32)優先日 平成13年7月17日(2001.7.17)  
(33)優先権主張国 日本 (JP)

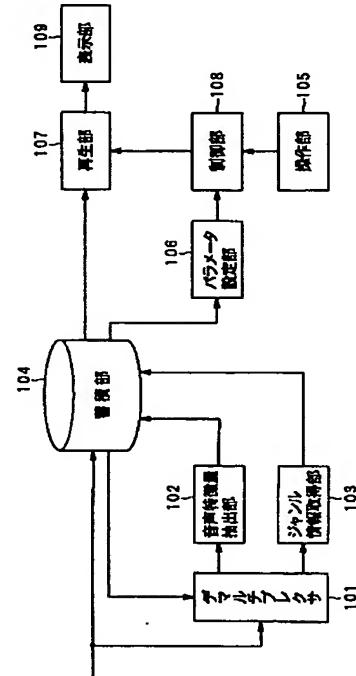
(71)出願人 000005016  
バイオニア株式会社  
東京都目黒区目黒1丁目4番1号  
(72)発明者 中村 譲  
埼玉県鶴ヶ島市富士見6丁目1番1号 バ  
イオニア株式会社総合研究所内  
(74)代理人 100083839  
弁理士 石川 泰男  
Fターム(参考) 5B075 ND12 ND14 NR12 NR16 NS01  
50052 AB04 AC08 DD04 DD06 EE03  
50053 FA20 FA21 HA30 JA01 JA30  
LA07 LA11  
5D015 HH00

(54)【発明の名称】 映像情報要約装置、映像情報要約方法及び映像情報要約プログラム

(57)【要約】

【課題】 映像情報の種別毎に的確に要約再生を行うことができるとともに、要約情報の生成処理の負担を軽減することのできる要約情報装置を提供すること。

【解決手段】 要約再生装置100は、入力された音声映像情報から予め設定されたパラメータに基づいて音声特徴量を取得する音声特徴量抽出部102と、入力された音声映像情報に付帯されている付帯情報からジャンル情報を取得するジャンル情報取得部103と、ジャンル情報に基づいて音声特徴量を抽出する際の最適なパラメータを設定する決定パラメータ設定部106と、設定されたパラメータに適合する音声特徴量に基づいて蓄積された音声映像情報における抽出すべき要約セグメントを決定するとともに、この要約セグメントに基づいて再生部107を制御する制御部108と、を備え、ジャンル毎に最適化されたパラメータによって要約再生を行うようになっている。



## 【特許請求の範囲】

【請求項1】 音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記コンテンツ情報の種別を識別する識別情報を取得する取得手段と、最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記コンテンツ情報を複数の前記コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えたことを特徴とする映像情報要約装置。

【請求項2】 音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記映像情報の種別を識別する識別情報を外部から取得する取得手段と、最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記音声情報を複数の前記音声区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えたことを特徴とする映像情報要約装置。

【請求項3】 請求項2に記載の映像情報要約装置において、前記決定手段が、前記分類した複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報を決定することを特徴とする映像情報要約装置。

【請求項4】 請求項2または3に記載の映像情報要約装置において、前記取得した識別情報によって無音部分を有する映像情報であることを識別した場合に、前記決定手段が、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間にに基づいて前記部分映像情報を決定するとともに、

前記最適化手段が、前記決定手段が前記無音区間を取得するときに用いられる前記閾値を最適化することを特徴とする映像情報要約装置。

【請求項5】 請求項2または3に記載の映像情報要約装置において、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歓聲音を有する映像情報であることを識別した場合に、

前記決定手段が、前記音声情報を分類することによって前記歓聲音を有する歓聲音区間を取得し、少なくとも当該歓聲音区間にに基づいて前記部分映像情報を決定するとともに、

前記最適化手段が、前記決定手段が前記歓聲音区間を取得するときに用いられる前記閾値を最適化することを特徴とする映像情報要約装置。

【請求項6】 音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、

前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、

前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段と、前記最適化された重要度を前記各部分映像情報に設定する設定手段と、

前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段と、

を備えたことを特徴とする映像情報要約装置。

【請求項7】 音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、

前記映像情報の種別を識別する識別情報を取得する取得手段と、

前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、

前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段と、前記最適化された重要度を前記各部分映像情報に設定する設定手段と、

前記決定された部分映像情報を前記重要度に基づいて前

3

記映像情報から抽出して前記要約情報を生成する生成手段と、

を備えたことを特徴とする映像情報要約装置。

【請求項8】 請求項7に記載の映像情報要約装置において、

前記決定手段が異なる閾値に基づいて前記抽出すべき部分映像情報を決定する場合に、

前記最適化手段が、前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度を最適化するとともに、

前記設定手段が、前記最適化された重要度を前記部分映像情報に設定することを特徴とする映像情報要約装置。

【請求項9】 請求項7または8に記載の映像情報要約装置において、

前記取得した識別情報によって無音部分を有する映像情報であること識別した場合に、

前記決定手段が、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間に基づいて前記部分映像情報を決定するとともに、

前記最適化手段が、前記無音区間に基づいて決定された前記部分映像情報に設定される前記重要度を最適化することを特徴とする映像情報要約装置。

【請求項10】 請求項7または8に記載の映像情報要約装置において、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歓声音を有する映像情報であることを識別した場合に、

前記決定手段が、前記音声情報を分類することによって前記歓声音を有する歓声音区間を取得し、少なくとも当該歓声音区間に基づいて前記部分映像情報を決定するとともに、

前記最適化手段が、前記歓声音区間に基づいて設定された前記部分映像情報に設定される前記重要度を最適化することを特徴とする映像情報要約装置。

【請求項11】 音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、

前記コンテンツ情報の種別を識別する識別情報を取得する取得処理工程と、

最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、

前記コンテンツ情報を複数の前記コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化処理工程と、

前記決定された前記部分映像情報を前記映像情報から抽

4

出して前記要約情報を生成する生成処理工程と、を備えたことを特徴とする映像情報要約方法。

【請求項12】 音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記映像情報の種別を識別する識別情報を外部から取得する取得処理工程と、

10 最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間に基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、

前記音声情報を複数の前記音声区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化処理工程と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を含むことを特徴とする映像情報要約方法。

20 【請求項13】 請求項12に記載の映像情報要約方法において、

前記決定処理工程においては前記分類された複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報が決定されることを特徴とする映像情報要約方法。

【請求項14】 請求項12または13に記載の映像情報要約方法において、

前記取得した識別情報によって無音部分を有する映像情報であることが識別された場合に、

30 前記決定処理工程においては前記音声情報を複数の前記音声区間に分類されるときに少なくとも一部に前記無音部分を有する無音区間が取得され、少なくとも当該無音区間に基づいて前記部分映像情報が決定されるとともに、

前記最適化処理工程においては前記無音区間が取得されるときに用いられる前記閾値が最適化されることを特徴とする映像情報要約方法。

40 【請求項15】 請求項13または14に記載の映像情報要約方法において、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歓声音を有する映像情報であることを識別された場合に、

前記決定処理工程においては前記音声情報を複数の前記音声区間に分類されるときに前記歓声音を有する歓声音区間が取得され、少なくとも当該歓声音区間に基づいて前記部分映像情報が決定されるとともに、

前記最適化処理工程においては前記歓声音区間が取得されるときに用いられる前記閾値が最適化されることを特徴とする映像情報要約方法。

【請求項16】 音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記コンテンツ情報の種別を識別する識別情報を取得する取得処理工程と、前記コンテンツ情報における閾値に基づいて前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化処理工程と、前記最適化された重要度を前記各部分映像情報に設定する設定処理工程手段と、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を備えたことを特徴とする映像情報要約方法。

【請求項17】 音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、

前記映像情報の種別を識別する識別情報を取得する取得処理工程と、前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間にに基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、

前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化処理工程と、

前記最適化された重要度を前記各部分映像情報に設定する設定処理工程と、

前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成処理工程と、

を含むことを特徴とする映像情報要約方法。

【請求項18】 請求項17に記載の映像情報要約方法において、

前記決定処理工程によって異なる閾値に基づいて前記抽出すべき部分映像情報が決定される場合に、

前記最適化処理工程においては前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度が最適化されるとともに、

前記設定処理工程においては前記最適化された重要度が前記部分映像情報に設定されることを特徴とする映像情報要約方法。

【請求項19】 請求項17または18に記載の映像情報要約方法において、

前記取得した識別情報によって無音部分を有する映像情報であることが識別された場合に、

前記決定処理工程においては前記音声情報が複数の前記音声区間に分類されるときに少なくとも一部に前記無音部分を有する無音区間が取得され、少なくとも当該無音区間にに基づいて前記部分映像情報が決定されるとともに、

10 前記最適化処理工程においては前記無音区間にに基づいて決定された前記部分映像情報に設定される前記重要度が最適化されることを特徴とする映像情報要約方法。

【請求項20】 請求項17または18に記載の映像情報要約方法において、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歎聲音を有する映像情報であることが識別された場合に、

前記決定処理工程においては前記音声情報が複数の前記音声区間に分類されたときに前記歎聲音を有する歎聲音区間が取得され、少なくとも当該歎聲音区間にに基づいて前記部分映像情報が決定されるとともに、

20 前記最適化処理工程においては前記歎聲音区間にに基づいて設定された前記部分映像情報に設定される前記重要度が最適化されることを特徴とする映像情報要約方法。

【請求項21】 コンピュータによって、音声情報及び映像情報からなるコンテンツ情報を、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、

前記コンピュータを、

前記コンテンツ情報を種別を識別する識別情報を取得する取得手段、

最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段、

前記コンテンツ情報を複数の前記コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段、及び、

40 前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段、として機能させることを特徴とする映像情報要約プログラム。

【請求項22】 コンピュータによって、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報

要約プログラムにおいて、  
前記コンピュータを、  
前記映像情報の種別を識別する識別情報を外部から取得する取得手段、  
最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間に基づいて前記抽出すべき部分映像情報を決定する決定手段、  
前記音声情報を複数の前記音声区間に分類するときに用いられる1または複数の閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段、  
前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項23】 請求項22に記載の映像情報要約プログラムにおいて、

前記コンピュータを、  
前記分類した複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報を決定する決定手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項24】 請求項22または23に記載の映像情報要約プログラムにおいて、

前記取得した識別情報によって無音部分を有する映像情報であることを識別した場合に、  
前記コンピュータを、

前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間に基づいて前記部分映像情報を決定する前記決定手段、  
前記決定手段が前記無音区間を取得するときに用いられる前記閾値を最適化する前記最適化手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項25】 請求項22または23に記載の映像情報要約プログラムにおいて、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歓聲音を有する映像情報であることを識別した場合に、  
前記コンピュータを、  
前記音声情報を分類することによって前記歓聲音を有する歓聲音区間を取得し、少なくとも当該歓聲音区間に基づいて前記部分映像情報を決定する決定手段、  
前記決定手段が前記歓聲音区間を取得するときに用いられる前記閾値を最適化する最適化手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項26】 コンピュータによって、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、  
前記コンピュータを、  
前記コンテンツ情報の種別を識別する識別情報を取得する取得手段、

10 前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて前記抽出すべき部分映像情報を決定する決定手段、  
前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段、  
前記最適化された重要度を前記各部分映像情報に設定する設定手段、及び、  
前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項27】 コンピュータによって、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、  
前記コンピュータを、  
前記映像情報の種別を識別する識別情報を取得する取得手段、前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間に基づいて前記抽出すべき部分映像情報を決定する決定手段と、  
前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段、  
前記最適化された重要度を前記各部分映像情報に設定する設定手段、  
前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段、  
として機能させることを特徴とする映像情報要約プログラム。

【請求項28】 請求項27に記載の映像情報要約プログラムにおいて、  
異なる閾値に基づいて前記抽出すべき部分映像情報を決定する場合に、  
前記コンピュータを、  
前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度を最適化する最適化手段、

前記最適化された重要度を前記部分映像情報に設定する設定手段、

として機能させることを特徴とする映像情報要約プログラム。

【請求項29】 請求項27または28に記載の映像情報要約プログラムにおいて、

前記取得した識別情報によって無音部分を有する映像情報であること識別した場合に、

前記コンピュータを、

前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間に基づいて前記部分映像情報を決定する決定手段、

前記無音区間に基づいて決定された前記部分映像情報に設定される前記重要度を最適化する最適化手段、

として機能させることを特徴とする映像情報要約プログラム。

【請求項30】 請求項27または28に記載の映像情報要約プログラムにおいて、

前記取得した識別情報によって前記音声情報を構成する背景雑音に歓聲音を有する映像情報であることを識別した場合に、

前記コンピュータを、

前記音声情報を分類することによって前記歓聲音を有する歓聲音区間を取得し、少なくとも当該歓聲音区間に基づいて前記部分映像情報を決定する決定手段、

前記歓聲音区間に基づいて設定された前記部分映像情報に設定される前記重要度を最適化する最適化手段、

として機能させることを特徴とする映像情報要約プログラム。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】 本発明は、音声が付帯された映像情報における要約再生装置の技術分野に属し、より詳細には、音声レベルに基づいて抽出すべき部分映像情報を決定する技術分野に属する。

##### 【0002】

【従来の技術】 近年、テレビジョン放送による番組など映像情報を録画・再生するVTR(Video Tape Recorder)などの記録装置が普及するにつれて、記録された映像情報を全て見ることなく当該映像情報を短時間に要約した映像情報(以下、要約情報という)を提供するダイジェスト再生(要約再生)が実用に供されるようになっている。

【0003】 一方、映像情報には、さまざまなもののが考えられ、例えば、テレビジョン放送によって提供される番組では、映画、ニュース番組、歌番組、スポーツ観戦番組などそのジャンルは多岐に亘る。このため、映像情報の種類によって短時間に要約する際の抽出すべき部分映像情報の特徴部分が異なるので、的確に映像情報の特

徴部分を抽出するためには映像情報の種類の情報を取得する必要がある。

【0004】 すなわち、スポーツ番組であれば、盛り上がった部分がその映像情報の特徴的な部分になるので、この盛り上がった部分を的確に抽出する必要があり、また、ニュース番組であれば、各ニュース毎の冒頭の部分には次のニュースの概略が示されるので、各ニュース内容が切り替わった部分を的確に抽出する必要がある。

【0005】 従来、このような映像種別を識別して要約再生を行う代表的なものとしては、特開平9-219835号公報記載の映像情報要約装置が知られている。

【0006】 同公報記載の映像情報要約装置1は、図7に示すように、入力された映像情報をデジタル化または復号化する映像情報入力部2と、映像情報入力部2から出力された映像情報から映像情報の種別を識別するジャンル情報取得部3と、映像種別を属性として映像区間の管理を行う映像区間管理部4と、映像種別に応じて映像情報およびその一部の区間の代表画素を作成する画素作成部5と、画素作成部5によって作成された各代表画素を空間的および時間的にレイアウトして要約映像情報を生成するレイアウト部6と、生成された映像情報を表示する表示部7とを備え、映像情報の種別毎に効率的な要約映像情報を生成できるようになっている。

##### 【0007】

【発明が解決しようとする課題】 しかしながら、上述のような要約再生方法にあっては、要約映像情報を生成する際に、映像情報の種別毎に異なる処理を行わなければならぬという問題を有していた。すなわち、上述の要約再生方法にあっては、ニュース番組、映画などの映像情報の種別毎に要約再生時に必要とされる代表画素を異なる手続によって作成し、要約映像情報を生成するので、要約映像情報の生成過程による処理の負担が増大し、また、この生成処理が煩雑になっていた。

【0008】 本発明は、上記の各問題点に鑑みて為されたもので、その課題は、映像情報の種別毎に的確に要約再生を行うことができるとともに、要約情報の生成処理の負担を軽減することのできる要約情報装置を提供することにある。

##### 【0009】

【課題を解決するための手段】 上記の課題を解決するために、請求項1に記載の発明は、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記コンテンツ情報の種別を識別する識別情報を取得する取得手段と、最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記コンテンツ情報を複数の前記

コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えた構成をしている。

【0010】この構成により、請求項1に記載の発明では、取得手段がコンテンツ情報の種別を識別する識別情報を取得し、最適化手段がこの識別情報に基づいて映像情報における1または複数の閾値を最適な値に設定すると、決定手段が最適化された閾値によってコンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて抽出すべき部分映像情報を決定するとともに、生成手段が決定された部分映像情報によって要約情報を生成する。

【0011】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報（映像情報を少なくとも含むコンテンツ情報であり、当該映像情報の他に音声情報又はデータ情報を含む場合もある。以下、同様）に含まれる映像情報の特徴は変化する。

【0012】例えば、ニュース番組には、各ニュースの合間にはいわゆるシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすことも多く、コンテンツ情報の特徴的な部分を示す。

【0013】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分では、シーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0014】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0015】したがって、本発明では、部分映像情報を決定するときに基準となる複数のコンテンツ区間に分類する際の閾値を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0016】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0017】上記の課題を解決するために、請求項2に記載の発明は、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部

分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記映像情報の種別を識別する識別情報を外部から取得する取得手段と、最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間に基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記音声情報を複数の前記音声区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えた構成を有している。

【0018】この構成により、請求項2に記載の発明では、取得手段が映像情報の種別を識別する識別情報を取得し、最適化手段がこの識別情報に基づいて音声情報を取得する1または複数の閾値を最適な値に設定すると、決定手段が最適化された閾値によって映像情報を複数の音声区間に分類し、当該分類した音声区間に基づいて抽出すべき部分映像情報を決定するとともに、生成手段が決定された部分映像情報によって要約情報を生成する。

【0019】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0020】例えば、ニュース番組には、各ニュースの合間には無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすことも多く、映像情報の特徴的な部分を示す。

【0021】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歓声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歓声の音声レベルが非常に大きくなるので、歓声の音声レベルが映像情報の特徴的な部分を示す。また、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歓声を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要がある。

【0022】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0023】したがって、本発明では、部分映像情報を決定するときに基準となる複数の音声区間に分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0024】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

13

【0025】また、請求項3に記載の発明は、請求項2に記載の映像情報要約装置において、前記決定手段が、前記分類した複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報を決定する構成を有している。

【0026】この構成により、請求項3に記載の発明では、決定手段が、分類した複数の音声区間の少なくとも何れか一種類の、映像情報中における少なくとも時間軸上の位置に基づいて抽出されるべき部分映像情報を決定する。

【0027】通常、映像情報に付帯される音声情報は、番組の盛り上がった部分、場面が転換された部分、または、内容が切り替わった部分などの特徴部分を示すので、当該映像情報の時間長を短縮して要約を行う際には重要な役割を果たす。

【0028】例えば、音声情報が付帯された映像情報において無音部分に続く部分は、次の内容の冒頭部分、特に当該内容の概略が示されることも多いので、無音部分の時間軸上の終了位置が映像情報の内容を把握する上で重要な部分となる。また、映像情報における盛り上がった音声が大きくなった騒音部分も特徴部分となるので、騒音部分の開始位置は映像情報の内容を把握する上で重要な部分となる。

【0029】したがって、本発明では、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0030】また、請求項4に記載の発明は、請求項2または3に記載の映像情報要約装置において、前記取得した識別情報によって無音部分を有する映像情報であることを識別した場合に、前記決定手段が、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間にに基づいて前記部分映像情報を決定するとともに、前記最適化手段が、前記決定手段が前記無音区間を取得するときに用いられる前記閾値を最適化する構成を有している。

【0031】この構成により、請求項4に記載の発明では、取得手段が取得した識別情報が無音部分を有する映像情報であることを識別した場合に、最適化手段が、決定手段が無音区間を取得するときに用いられる閾値を最適化するとともに、決定手段がこの最適化された閾値に基づいて音声情報から無音区間を取得して抽出すべき部分映像情報を決定する。

【0032】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像

14

情報では、特徴部分を抽出する際に無音部分を的確に検出することが重要となる。

【0033】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取得することができるので、映像情報の内容に即した要約情報を得ることができる。

【0034】また、請求項5に記載の発明は、請求項2または3に記載の映像情報要約装置において、前記取得した識別情報によって前記音声情報を構成する背景雑音に歎声音を有する映像情報であることを識別した場合に、前記決定手段が、前記音声情報を分類することによって前記歎声音を有する歎声音区間を取得し、少なくとも当該歎声音区間にに基づいて前記部分映像情報を決定するとともに、前記最適化手段が、前記決定手段が前記歎声音区間を取得するときに用いられる前記閾値を最適化する構成を有している。

【0035】この構成により、請求項5に記載の発明20は、取得手段が取得した映像情報の識別情報が音声情報を構成する背景雑音に歎声音を有する映像情報であることを識別した場合に、最適化手段が、決定手段が歎声音区間を取得するときの閾値を最適化するとともに、決定手段がこの最適化された閾値に基づいて音声情報から無音区間を取得して抽出すべき部分映像情報を決定する。

【0036】映像情報における音声特性は当該映像情報の種別によって異なるとともに、当該映像情報に付帯される音声情報を構成する背景雑音に歎声音を有する映像情報では、特徴部分を抽出する際に歎声音区間を的確に検出することが重要となる。

【0037】したがって、本発明では、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歎声音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歎声音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができる。

【0038】また、請求項6に記載された発明は、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記コンテンツ情報の種別を識別する識別情報を取得する取得手段と、前記コンテンツ情報における閾値に基づいて前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段と、前記最適化された重要度を前記各部分映像情報に設定する設定手段と、前記決定された部分映像情

50

報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えた構成をしている。

【0039】この構成により、請求項6に記載の発明では、取得手段がコンテンツ情報の種別を識別する識別情報を取得し、決定手段が閾値に基づいてコンテンツ情報を複数のコンテンツ区間に分類して抽出すべき部分映像情報を決定するとともに、最適化手段が各部分映像情報に設定される重要度を、取得した識別情報に基づいて最適化すると、設定手段が最適化された重要度を部分映像情報に設定し、生成手段が決定された部分映像情報および重要度に基づいて要約情報を生成する。

【0040】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報に付帯される映像情報の特徴は変化する。

【0041】例えば、ニュース番組には、各ニュースの合間にはシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、コンテンツ情報の特徴的な部分を示す。したがって、シーンチェンジ以外の他の区間に比べ重要度は高い。

【0042】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分ではシーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0043】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0044】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができる。コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0045】また、請求項7に記載された発明は、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約装置において、前記映像情報の種別を識別する識別情報を取得する取得手段と、前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段と、前記最適化された重要度を前記各部分映像情報に設定する設定手段

と、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段と、を備えた構成を有している。

【0046】この構成により、請求項7に記載の発明では、取得手段が映像情報の種別を識別する識別情報を取得し、決定手段が閾値に基づいて映像情報を複数の音声区間に分類して抽出すべき部分映像情報を決定するとともに、最適化手段が各部分映像情報に設定される重要度を、取得した識別情報に基づいて最適化すると、設定手段が最適化された重要度を部分映像情報に設定し、生成手段が決定された部分映像情報および重要度に基づいて要約情報を生成する。

【0047】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0048】例えば、ニュース番組には、各ニュースの合間には無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、映像情報の特徴的な部分を示す。したがって、騒音区間などの他の音声区間に比べ重要度は高い。

【0049】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歎声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歎声の音声レベルが非常に大きくなるので、歎声の音声レベルが映像情報の特徴的な部分を示す。さらに、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歎聲音を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要があるとともに、盛り上がった内容を的確に抽出して要約再生を行う場合に、基準となる区間によって要約再生時の重要度の設定を変える必要がある。

【0050】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0051】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができる。映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0052】また、請求項8に記載の発明は、請求項7に記載の映像情報要約装置において、前記決定手段が異なる閾値に基づいて前記抽出すべき部分映像情報を決定する場合に、前記最適化手段が、前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度を最適化するとともに、前記設定手段が、前記最適化された重要度を前記部分映像情報に設定する構成を有している。

【0053】この構成により、請求項8に記載の発明で

は、決定手段が異なる閾値に基づいて抽出すべき部分映像情報を決定する場合に、設定手段が異なる閾値毎に取得した識別情報に基づいて最適化した重要度を、決定された各部分映像情報を設定する。

【0054】映像情報の種別によって映像情報に付帯される音声特性が異なり、要約再生が行われる映像情報毎に音声区間の重要度が異なる。特に、ニュース番組など無音区間を有する映像情報では、音声レベルが高い区間に比べ無音区間の重要度は高い。また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歓声音を有する映像情報では、無音区間より音声レベルが高い区間の重要度は高く、また、複数の音声レベルによって部分映像情報を決定する場合は、異なる音声レベルによって要約再生時の重要度が異なる。

【0055】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0056】また、請求項9に記載の発明は、請求項7または8に記載の映像情報要約装置において、前記取得した識別情報によって無音部分を有する映像情報であること識別した場合に、前記決定手段が、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間に基づいて前記部分映像情報を決定するとともに、前記最適化手段が、前記無音区間に基づいて決定された前記部分映像情報に設定される前記重要度を最適化する構成をしている。

【0057】この構成により、請求項9に記載の発明は、取得した識別情報によって無音部分を有する映像情報であること識別した場合に、決定手段が前記音声情報を分類することによって少なくとも一部に無音部分を有する無音区間を取得し、当該無音区間に基づいて部分映像情報を決定するとともに、最適化手段が無音区間に基づいて決定された前記部分映像情報を設定される前記重要度を最適化する。

【0058】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像情報では、要約再生する時に無音部分を的確に再生することが重要となる。

【0059】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情

報を得ることができる。

【0060】また、請求項10に記載の発明は、請求項7または8に記載の映像情報要約装置において、前記取得した識別情報によって前記音声情報を構成する背景雑音に歓声音を有する映像情報であることを識別した場合に、前記決定手段が、前記音声情報を分類することによって前記歓声音を有する歓声音区間を取得し、少なくとも当該歓声音区間に基づいて前記部分映像情報を決定するとともに、前記最適化手段が、前記歓声音区間に基づいて設定された前記部分映像情報に設定される前記重要度を最適化する構成を有している。

【0061】この構成により、請求項10に記載の発明は、取得した識別情報によって音声情報を構成する背景雑音に歓声音を有する映像情報であることを識別した場合に、決定手段が前記音声区間を分類することによって歓声音を有する歓声音区間を取得し、当該歓声音区間に基づいて部分映像情報を決定するとともに、最適化手段が歓声音区間に基づいて決定された部分映像情報に設定される重要度を最適化する。

【0062】映像情報における音声特性は当該映像情報の種別によって異なるとともに、音声情報を構成する背景雑音に歓声音を有する映像情報では、要約再生する時に大きな歓声音を有する区間を的確に再生することが重要となる。

【0063】したがって、本発明では、映像情報の識別情報により、音声情報を構成する背景雑音に歓声音を有する映像情報であることが識別された場合に、歓声音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0064】また、請求項11に記載の発明は、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記コンテンツ情報の種別を識別する識別情報を取得する取得処理工程と、最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、前記コンテンツ情報を複数の前記コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化処理工程と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を含む構成を有している。

【0065】この構成により、請求項11に記載の発明では、取得処理工程によってコンテンツ情報の種別を識別する識別情報を取得し、最適化処理工程によってこの

19

識別情報に基づいて映像情報における 1 または複数の閾値を最適な値に設定すると、決定処理工程において最適化された閾値によってコンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて抽出すべき部分映像情報を決定するとともに、生成処理工程が決定された部分映像情報によって要約情報を生成する。

【0066】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報に含まれる映像情報の特徴は変化する。

【0067】例えば、ニュース番組には、各ニュースの合間にはいわゆるシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすことも多く、コンテンツ情報の特徴的な部分を示す。

【0068】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分では、シーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0069】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0070】したがって、本発明では、部分映像情報を決定するときに基準となる複数のコンテンツ区間に分類する際の閾値を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0071】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0072】また、請求項12に記載の発明は、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を 1 または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記映像情報の種別を識別する識別情報を外部から取得する取得処理工程と、最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間にに基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、前記音声情報を複数の前記音声区間に分類するときに用いられる 1 または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化処

20

理工程と、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を含む構成を有している。

【0073】この構成により、請求項12に記載の発明では、取得処理工程によって映像情報の種別を識別する識別情報が取得され、最適化処理工程によってこの識別情報に基づいて音声情報における 1 または複数の閾値が最適な値に設定されると、決定処理工程によって最適化された閾値により映像情報が複数の音声区間に分類され、当該分類された音声区間にに基づいて抽出すべき部分映像情報が決定されるとともに、生成処理工程によって決定された部分映像情報により要約情報が生成される。

【0074】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0075】例えば、ニュース番組には、各ニュースの合間には無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすことも多く、映像情報の特徴的な部分を示す。

【0076】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歓声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歓声の音声レベルが非常に大きくなるので、歓声の音声レベルが映像情報の特徴的な部分を示す。また、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歓声を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要がある。

【0077】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0078】したがって、本発明では、部分映像情報を決定するときに基準となる複数の音声区間に分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0079】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0080】また、請求項13に記載の発明は、請求項12に記載の映像情報要約方法において、前記決定処理工程においては前記分類された複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報が決定される構成を有している。

【0081】この構成により、請求項13に記載の発明では、決定処理工程によって分類された複数の音声区間

21

の少なくとも何れか一種類の、映像情報中における少なくとも時間軸上の位置に基づいて抽出されるべき部分映像情報を決定する。

【0082】通常、映像情報に付帯される音声情報は、番組の盛り上がった部分、場面が転換された部分、または、内容が切り替わった部分などの特徴部分を示すので、当該映像情報の時間長を短縮して要約を行う際には重要な役割を果たす。

【0083】例えば、音声情報が付帯された映像情報において無音部分に続く部分は、次の内容の冒頭部分、特に当該内容の概略が示されることも多いので、無音部分の時間軸上の終了位置が映像情報の内容を把握する上で重要な部分となる。また、映像情報における盛り上がった音声が大きくなったり騒音部分も特徴部分となるので、騒音部分の開始位置は映像情報の内容を把握する上で重要な部分となる。

【0084】したがって、本発明では、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0085】また、請求項14に記載の発明は、請求項12または13に記載の映像情報要約方法において、前記取得した識別情報によって無音部分を有する映像情報であることが識別された場合に、前記決定処理工程においては前記音声情報が複数の前記音声区間に分類されるときに少なくとも一部に前記無音部分を有する無音区間が取得され、少なくとも当該無音区間にに基づいて前記部分映像情報が決定されるとともに、前記最適化処理工程においては前記無音区間が取得されるときに用いられる前記閾値が最適化される構成を有している。

【0086】この構成により、請求項14に記載の発明では、取得した識別情報が無音部分を有する映像情報であることが識別された場合に、最適化処理工程によって無音区間を取得するときに用いられる閾値が最適化されるとともに、決定処理工程によってこの最適化された閾値に基づいて音声情報から無音区間が取得されて抽出すべき部分映像情報が決定される。

【0087】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像情報では、特徴部分を抽出する際に無音部分を的確に検出することが重要となる。

【0088】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取得することができるので、映像情報の内容に即した要約情報を得ることができる。

22

【0089】また、請求項15に記載の発明は、請求項13または14に記載の映像情報要約方法において、前記取得した識別情報によって前記音声情報を構成する背景雑音に歎聲音を有する映像情報であることが識別された場合に、前記決定処理工程においては前記音声情報が複数の前記音声区間に分類されたときに前記歎聲音を有する歎聲音区間が取得され、少なくとも当該歎聲音区間にに基づいて前記部分映像情報が決定されるとともに、前記最適化処理工程においては前記歎聲音区間が取得されたときに用いられる前記閾値が最適化される構成を有している。

【0090】この構成により、請求項15に記載の発明は、取得した映像情報の識別情報が音声情報を構成する背景雑音に歎聲音を有する映像情報であることが識別された場合に、最適化処理工程によって歎聲音区間が取得されるときの閾値が最適化されるとともに、決定処理工程によってこの最適化された閾値に基づいて音声情報から無音区間が取得されて抽出すべき部分映像情報が決定される。

【0091】映像情報における音声特性は当該映像情報の種別によって異なるとともに、当該映像情報に付帯される音声情報を構成する背景雑音に歎聲音を有する映像情報では、特徴部分を抽出する際に歎聲音区間に的確に検出することが重要となる。

【0092】したがって、本発明では、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歎聲音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歎聲音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができる。

【0093】また、請求項16に記載の発明は、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記コンテンツ情報の種別を識別する識別情報を取得する取得処理工程と、前記コンテンツ情報における閾値に基づいて前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化処理工程と、前記最適化された重要度を前記各部分映像情報に設定する設定処理工程手段と、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を含む構成を有している。

【0094】この構成により、請求項16に記載の発明では、取得処理工程によってコンテンツ情報の種別を識別する識別情報を取得し、決定処理工程によって閾値に

23

基づいてコンテンツ情報を複数のコンテンツ区間に分類して抽出すべき部分映像情報を決定するとともに、最適化処理工程によって各部分映像情報に設定される重要度を、取得した識別情報に基づいて最適化すると、設定処理工程によって最適化された重要度を部分映像情報に設定し、生成処理工程によって決定された部分映像情報および重要度に基づいて要約情報を生成する。

【0095】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報に付帯される映像情報の特徴は変化する。

【0096】例えば、ニュース番組には、各ニュースの合間にはシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、コンテンツ情報の特徴的な部分を示す。したがって、シーンチェンジ以外の他の区間に比べ重要度は高い。

【0097】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分ではシーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0098】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0099】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができる、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0100】また、請求項17に記載の発明は、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約方法において、前記映像情報の種別を識別する識別情報を取得する取得処理工程と、前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間にに基づいて前記抽出すべき部分映像情報を決定する決定処理工程と、前記各部分映像情報に設定される重要度を、前記取得した識別情報に基づいて最適化する最適化処理工程と、前記最適化された重要度を前記各部分映像情報に設定する設定処理工程と、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成処理工程と、を含む構成を有している。

【0101】この構成により、請求項17に記載の発明では、取得処理工程によって映像情報の種別を識別する

24

識別情報が取得され、決定処理工程によって閾値に基づいて映像情報が複数の音声区間に分類されて抽出すべき部分映像情報が決定されるとともに、最適化処理工程によって各部分映像情報に設定される重要度が、取得された識別情報に基づいて最適化されると、設定処理工程によって最適化された重要度が部分映像情報に設定され、生成処理工程によって決定された部分映像情報および重要度に基づいて要約情報が生成される。

【0102】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0103】例えば、ニュース番組には、各ニュースの合間には無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、映像情報の特徴的な部分を示す。したがって、騒音区間などの他の音声区間に比べ重要度は高い。

【0104】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歓声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歓声の音声レベルが非常に大きくなるので、歓声の音声レベルが映像情報の特徴的な部分を示す。さらに、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歓声を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要があるとともに、盛り上がった内容を的確に抽出して要約再生を行う場合に、基準となる区間によって要約再生時の重要度の設定を変える必要がある。

【0105】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0106】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0107】また、請求項18に記載の発明は、請求項17に記載の映像情報要約方法において、前記決定処理工程によって異なる閾値に基づいて前記抽出すべき部分映像情報が決定される場合に、前記最適化処理工程においては前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度が最適化されるとともに、前記設定処理工程においては前記最適化された重要度が前記部分映像情報に設定される構成を有している。

【0108】この構成により、請求項18に記載の発明では、決定処理工程によって異なる閾値に基づいて抽出すべき部分映像情報が決定される場合に、設定処理工程によって異なる閾値毎に取得した識別情報に基づいて最

25

適化された重要度が、決定された各部分映像情報に設定される。

【0109】映像情報の種別によって映像情報に付帯される音声特性が異なり、要約再生が行われる映像情報毎に音声区間の重要度が異なる。特に、ニュース番組など無音区間を有する映像情報では、音声レベルが高い区間に比べ無音区間の重要度は高い。また、スポーツ観戦番組などの該映像情報に付帯される音声情報を構成する背景雑音に歓声音を有する映像情報では、無音区間より音声レベルが高い区間の重要度は高く、また、複数の音声レベルによって部分映像情報を決定する場合は、異なる音声レベルによって要約再生時の重要度が異なる。

【0110】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0111】また、請求項19に記載の発明は、請求項17または18に記載の映像情報要約方法において、前記取得した識別情報によって無音部分を有する映像情報であることが識別された場合に、前記決定処理工程においては前記音声情報が複数の前記音声区間に分類されるときに少なくとも一部に前記無音部分を有する無音区間が取得され、少なくとも当該無音区間に基づいて前記部分映像情報が決定されるとともに、前記最適化処理工程によって前記無音区間に基づいて決定された前記部分映像情報に設定される前記重要度が最適化される構成を有している。

【0112】この構成により、請求項19に記載の発明は、取得した識別情報によって無音部分を有する映像情報であることが識別された場合に、決定処理工程によって分類された音声区間の中から少なくとも一部に無音部分を有する無音区間が取得され、当該無音区間に基づいて部分映像情報が決定されるとともに、最適化処理工程によって無音区間に基づいて決定された前記部分映像情報に設定される前記重要度が最適化される。

【0113】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像情報では、要約再生する時に無音部分を的確に再生することが重要となる。

【0114】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

26

【0115】また、請求項20に記載の発明は、請求項17または18に記載の映像情報要約方法において、前記取得した識別情報によって前記音声情報を構成する背景雑音に歓声音を有する映像情報であることが識別された場合に、前記決定処理工程によって前記音声情報が複数の前記音声区間に分類されるときに前記歓声音を有する歓声音区間が取得され、少なくとも当該歓声音区間に基づいて前記部分映像情報が決定されるとともに、前記最適化処理工程によって前記歓声音区間に基づいて設定された前記部分映像情報に設定される前記重要度が最適化される構成を有している。

【0116】この構成により、請求項20に記載の発明は、取得した識別情報によって音声情報を構成する背景雑音に歓声音を有する映像情報であることが識別された場合に、決定処理工程によって前記音声情報が複数の前記音声区間に分類されるときに歓声音を有する歓声音区間が取得され、当該歓声音区間に基づいて部分映像情報が決定されるとともに、最適化処理工程によって歓声音区間に基づいて決定された部分映像情報に設定される重要度が最適化される。

【0117】映像情報における音声特性は当該映像情報の種別によって異なるとともに、音声情報を構成する背景雑音に歓声音を有する映像情報では、要約再生する時に大きな歓声音を有する区間を的確に再生することが重要となる。

【0118】したがって、本発明では、映像情報の識別情報により、音声情報を構成する背景雑音に歓声音を有する映像情報であることが識別された場合に、歓声音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0119】また、請求項21に記載の発明は、コンピュータによって、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、前記コンピュータを、前記コンテンツ情報の種別を識別する識別情報を取得する取得手段、最適化された閾値によって前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間に基づいて前記抽出すべき部分映像情報を決定する決定手段、前記コンテンツ情報を複数の前記コンテンツ区間に分類するときに用いられる1または複数の前記閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段、及び、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段、として機能させる構成を有している。

【0120】この構成により、請求項21に記載の発明

27

では、コンピュータによって、コンテンツ情報の種別を識別する識別情報を取得し、この識別情報に基づいて映像情報における1または複数の閾値を最適な値に設定すると、最適化された閾値によってコンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて抽出すべき部分映像情報を決定するとともに、決定された部分コンテンツ情報によって要約情報を生成する。

【0121】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報に含まれる映像情報の特徴は変化する。

【0122】例えば、ニュース番組には、各ニュースの合間にはいわゆるシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすことも多く、コンテンツ情報の特徴的な部分を示す。

【0123】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分では、シーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0124】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0125】したがって、本発明では、部分コンテンツ情報を決定するときに基準となる複数のコンテンツ区間を分類する際の閾値を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0126】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0127】また、請求項22に記載の発明は、コンピュータによって、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、前記コンピュータを、前記映像情報の種別を識別する識別情報を外部から取得する取得手段、前記最適化された閾値によって前記映像情報に付帯されている音声情報を複数の音声区間に分類し、当該分類した音声区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段、前記音声情報を複数の前記音声区間に分類するとき

28

に用いられる1または複数の閾値を前記取得した識別情報に基づいて最適な値に設定する最適化手段、前記決定された前記部分映像情報を前記映像情報から抽出して前記要約情報を生成する生成手段、として機能させる構成を有している。

【0128】この構成により、請求項22に記載の発明では、コンピュータによって、映像情報の種別を識別する識別情報を取得し、この識別情報に基づいて音声情報における1または複数の閾値を最適な値に設定すると、

10 最適化された閾値によって映像情報を複数の音声区間に分類し、当該分類した音声区間にに基づいて抽出すべき部分映像情報を決定するとともに、決定された部分映像情報によって要約情報を生成する。

【0129】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0130】例えば、ニュース番組には、各ニュースの合間には無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示めすこと多く、映像情報の特徴的な部分を示す。

【0131】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歓声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歓声の音声レベルが非常に大きくなるので、歓声の音声レベルが映像情報の特徴的な部分を示す。また、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歓声を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要がある。

【0132】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0133】したがって、本発明では、部分映像情報を決定するときに基準となる複数の音声区間を分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

40 【0134】また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0135】また、請求項23に記載の発明は、請求項22に記載の映像情報要約プログラムにおいて、前記コンピュータを、前記分類した複数の音声区間の少なくとも何れか一種類の、前記映像情報中における少なくとも時間軸上の位置に基づいて前記抽出すべき前記部分映像情報を決定する決定手段、として機能させる構成を有している。

29

【0136】この構成により、請求項23に記載の発明では、コンピュータによって、分類した複数の音声区間の少なくとも何れか一種類の、映像情報中における少なくとも時間軸上の位置に基づいて抽出されるべき部分映像情報を決定する。

【0137】通常、映像情報に付帯される音声情報は、番組の盛り上がった部分、場面が転換された部分、または、内容が切り替わった部分などの特徴部分を示すので、当該映像情報の時間長を短縮して要約を行う際には重要な役割を果たす。

【0138】例えば、音声情報が付帯された映像情報において無音部分に続く部分は、次の内容の冒頭部分、特に当該内容の概略が示されることも多いので、無音部分の時間軸上の終了位置が映像情報の内容を把握する上で重要な部分となる。また、映像情報における盛り上がった音声が大きくなったり騒音部分も特徴部分となるので、騒音部分の開始位置は映像情報の内容を把握する上で重要な部分となる。

【0139】したがって、本発明では、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができます。

【0140】また、請求項24に記載の発明は、請求項22または23に記載の映像情報要約プログラムにおいて、前記取得した識別情報によって無音部分を有する映像情報であることを識別した場合に、前記コンピュータを、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間にに基づいて前記部分映像情報を決定する前記決定手段、前記決定手段が前記無音区間を取得するときに用いられる前記閾値を最適化する前記最適化手段、として機能させる構成を有している。

【0141】この構成により、請求項24に記載の発明では、取得した識別情報が無音部分を有する映像情報であることを識別した場合に、コンピュータによって、無音区間を取得するときに用いられる閾値を最適化するとともに、この最適化された閾値に基づいて音声情報から無音区間を取得して抽出すべき部分映像情報を決定する。

【0142】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像情報では、特徴部分を抽出する際に無音部分を的確に検出することが重要となる。

【0143】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取

30

得することができるので、映像情報の内容に即した要約情報を得ることができます。

【0144】また、請求項25に記載の発明は、請求項22または23に記載の映像情報要約プログラムにおいて、前記取得した識別情報によって前記音声情報を構成する背景雑音に歎聲音を有する映像情報であることを識別した場合に、前記コンピュータを、前記音声情報を分類することによって前記歎聲音を有する歎聲音区間を取得し、少なくとも当該歎聲音区間にに基づいて前記部分映像情報を決定する決定手段、前記決定手段が前記歎聲音区間を取得するときに用いられる前記閾値を最適化する最適化手段、として機能させる構成を有している。

【0145】この構成により、請求項25に記載の発明は、取得した映像情報の識別情報が音声情報を構成する背景雑音に歎聲音を有する映像情報であることを識別した場合に、コンピュータによって、歎聲音区間を取得するときの閾値を最適化するとともに、この最適化された閾値に基づいて音声情報から無音区間を取得して抽出すべき部分映像情報を決定する。

【0146】映像情報における音声特性は当該映像情報の種別によって異なるとともに、当該映像情報に付帯される音声情報を構成する背景雑音に歎聲音を有する映像情報では、特徴部分を抽出する際に歎聲音区間に的確に検出することが重要となる。

【0147】したがって、本発明では、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歎聲音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歎聲音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができます。

【0148】また、請求項26に記載の発明は、コンピュータによって、音声情報及び映像情報からなるコンテンツ情報から、前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、前記コンピュータを、前記コンテンツ情報の種別を識別する識別情報を取得する取得手段、前記コンテンツ情報における閾値に基づいて前記コンテンツ情報を複数のコンテンツ区間に分類し、当該分類したコンテンツ区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段、前記最適化された重要度を前記各部分映像情報に設定する設定手段、及び、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段、として機能させる構成を有している。

【0149】この構成により、請求項26に記載の発明では、コンピュータによって、コンテンツ情報の種別を

31

識別する識別情報を取得し、閾値に基づいてコンテンツ情報を複数のコンテンツ区間に分類して抽出すべき部分映像情報を決定するとともに、各部分映像情報に設定される重要度を、取得した識別情報に基づいて最適化すると、最適化された重要度を部分映像情報に設定し、決定された部分映像情報および重要度に基づいて要約情報を生成する。

【0150】通常、テレビジョン放送などの番組のジャンルによって、コンテンツ情報に付帯される映像情報の特徴は変化する。

【0151】例えば、ニュース番組には、各ニュースの合間にシーンチェンジ部分が存在する。すなわち、ニュース番組では、シーンチェンジ部分は、場面転換された部分または内容が切り替わった部分を示すとともに、このシーンチェンジ部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、コンテンツ情報の特徴的な部分を示す。したがって、シーンチェンジ以外の他の区間に比べ重要度は高い。

【0152】また、スポーツ観戦番組などの当該コンテンツ情報に付帯される映像情報において、内容が盛り上がった部分ではシーンチェンジが頻繁に行われる場合が多くなるので、当該シーンチェンジの間隔がコンテンツ情報の特徴的な部分を示す。

【0153】このように、コンテンツ情報の種別によって当該コンテンツ情報に含まれる映像情報の特性が異なる。

【0154】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0155】また、請求項27に記載の発明は、コンピュータによって、音声情報が付帯されている映像情報から、当該音声情報に基づいて前記映像情報の一部である部分映像情報を1または複数個抽出し、当該抽出した部分映像情報および重要度に基づいて前記映像情報の時間長が短縮された要約情報を生成する映像情報要約プログラムにおいて、前記コンピュータを、前記映像情報の種別を識別する識別情報を取得する取得手段、前記音声情報における閾値に基づいて前記映像情報を複数の音声区間に分類し、当該分類した区間にに基づいて前記抽出すべき部分映像情報を決定する決定手段と、前記各部分映像情報に設定される前記重要度を、前記取得した識別情報に基づいて最適化する最適化手段、前記最適化された重要度を前記各部分映像情報に設定する設定手段、前記決定された部分映像情報を前記重要度に基づいて前記映像情報から抽出して前記要約情報を生成する生成手段、として機能させる構成を有している。

【0156】この構成により、請求項27に記載の発明

32

では、コンピュータによって、映像情報の種別を識別する識別情報を取得し、閾値に基づいて映像情報を複数の音声区間に分類して抽出すべき部分映像情報を決定するとともに、各部分映像情報に設定される重要度を、取得した識別情報に基づいて最適化すると、最適化された重要度を部分映像情報に設定し、生成手段が決定された部分映像情報および重要度に基づいて要約情報を生成する。

【0157】通常、テレビジョン放送などの番組のジャンルによって、映像情報に付帯される音声情報の特徴は変化する。

【0158】例えば、ニュース番組には、各ニュースの合間に無音部分が存在する。すなわち、ニュース番組では、無音部分は、場面転換された部分または内容が切り替わった部分を示すとともに、この無音部分に続く部分は次の内容の冒頭部分、特に、当該内容の概略を示すことも多く、映像情報の特徴的な部分を示す。したがって、騒音区間などの他の音声区間に比べ重要度は高い。

【0159】また、スポーツ観戦番組などの当該映像情報に付帯される音声情報を構成する背景雑音に歎声を有する映像情報において、内容が盛り上がった部分では、音声情報に含まれる歎声の音声レベルが非常に大きくなるので、歎声の音声レベルが映像情報の特徴的な部分を示す。さらに、スポーツ番組には無音部分が存在しないまたは極端に少ない一方、常に背景雑音に歎声を有するため、内容の盛り上がりを示す音声区間の閾値が他の映像情報に比べ高く設定する必要があるとともに、盛り上がった内容を的確に抽出して要約再生を行う場合に、基準となる区間によって要約再生時の重要度の設定を変える必要がある。

【0160】このように、映像情報の種別によって当該映像情報に付帯される音声情報の音声特性が異なる。

【0161】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0162】また、請求項28に記載の発明は、請求項27に記載の映像情報要約プログラムにおいて、異なる閾値に基づいて前記抽出すべき部分映像情報を決定する場合に、前記コンピュータを、前記取得した識別情報に基づいて異なる前記閾値毎に前記重要度を最適化する最適化手段、前記最適化された重要度を前記部分映像情報に設定する設定手段、として機能させる構成を有している。

【0163】この構成により、請求項28に記載の発明では、異なる閾値に基づいて抽出すべき部分映像情報を決定する場合に、コンピュータによって、異なる閾値毎に取得した識別情報に基づいて最適化した重要度を、決

定された各部分映像情報に設定する。

【0164】映像情報の種別によって映像情報に付帯される音声特性が異なり、要約再生が行われる映像情報毎に音声区間の重要度が異なる。特に、ニュース番組など無音区間を有する映像情報では、音声レベルが高い区間に比べ無音区間の重要度は高い。また、スポーツ観戦番組などの該映像情報に付帯される音声情報を構成する背景雑音に歓聲音を有する映像情報では、無音区間より音声レベルが高い区間の重要度は高く、また、複数の音声レベルによって部分映像情報を決定する場合は、異なる音声レベルによって要約再生時の重要度が異なる。

【0165】したがって、本発明では、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0166】また、請求項29に記載の発明は、請求項27または28に記載の映像情報要約プログラムにおいて、前記取得した識別情報によって無音部分を有する映像情報であること識別した場合に、前記コンピュータを、前記音声情報を分類することによって少なくとも一部に前記無音部分を有する無音区間を取得し、少なくとも当該無音区間に基づいて前記部分映像情報を決定する決定手段、前記無音区間に基づいて決定された前記部分映像情報に設定される前記重要度を最適化する最適化手段、として機能させる構成を有している。

【0167】この構成により、請求項29に記載の発明は、取得した識別情報によって無音部分を有する映像情報であること識別した場合に、コンピュータにより、前記音声情報を分類することによって少なくとも一部に無音部分を有する無音区間を取得し、当該無音区間に基づいて部分映像情報を決定するとともに、無音区間に基づいて決定された前記部分映像情報に設定される前記重要度を最適化する。

【0168】映像情報における音声特性は当該映像情報の種別によって異なるとともに、無音部分を有する映像情報では、要約再生する時に無音部分を的確に再生することが重要となる。

【0169】したがって、本発明では、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0170】また、請求項30に記載の発明は、請求項27または28に記載の映像情報要約プログラムにおいて

て、前記取得した識別情報によって前記音声情報を構成する背景雑音に歓聲音を有する映像情報であることを識別した場合に、前記コンピュータを、前記音声情報を分類することによって前記歓聲音を有する歓聲音区間を取得し、少なくとも当該歓聲音区間に基づいて前記部分映像情報を決定する決定手段、前記歓聲音区間に基づいて設定された前記部分映像情報に設定される前記重要度を最適化する最適化手段、として機能させる構成を有している。

10 【0171】この構成により、請求項30に記載の発明は、取得した識別情報によって音声情報を構成する背景雑音に歓聲音を有する映像情報であることを識別した場合に、コンピュータにより、前記音声区間を分類することによって歓聲音を有する歓聲音区間を取得し、当該歓聲音区間に基づいて部分映像情報を決定するとともに、歓聲音区間に基づいて決定された部分映像情報に設定される重要度を最適化する。

【0172】映像情報における音声特性は当該映像情報の種別によって異なるとともに、音声情報を構成する背景雑音に歓聲音を有する映像情報では、要約再生する時に大きな歓聲音を有する区間を的確に再生することが重要となる。

【0173】したがって、本発明では、映像情報の識別情報により、音声情報を構成する背景雑音に歓聲音を有する映像情報であることが識別された場合に、歓聲音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

30 【0174】

【発明の実施の形態】 (I) 実施形態

次に、本発明に好適な実施の形態について、図面に基づいて説明する。

【0175】本実施形態は、通信回線または電波などによって提供されるテレビジョン放送の番組などの音声映像情報を要約再生する要約再生装置に対して本発明を適用した場合の実施形態である。

【0176】まず、図1を用いて本実施形態における要約再生装置の全体の構成および概要動作について説明する。

40 【0177】なお、図1は本実施形態に係る要約再生装置の構成を示すブロック図である。

【0178】図1に示す本実施形態の要約再生装置100は、通信回線から送信された、または、図示しない受信部によって受信されたデジタル音声映像情報が入力されるようになっており、この入力されたデジタル音声映像情報から当該音声映像情報の種別情報（以下、ジャンル情報という）を取得するようになっている。

【0179】また、この要約再生装置100は、予め設定された複数の閾値（以下、単にパラメータという）に

基づいて入力された音声映像情報の特徴量（以下、音声特徴量という）を抽出するとともに、入力された音声映像情報から取得したジャンル情報に基づいて基準となる閾値（以下、決定パラメータという）を設定するようになっており、当該設定した決定パラメータに適合するパラメータによって抽出した音声特徴量を選択するとともに、当該選択された音声特徴量に基づいて要約再生に際して抽出すべき部分映像情報（以下、要約セグメントという）を決定（以下、要約セグメントの決定処理という）するようになっている。

【0180】このように、この要約再生装置100は、抽出すべき要約セグメントを決定し、決定された要約セグメントに基づいて音声映像情報から要約セグメントを抽出して要約再生を行うようになっている。

【0181】また、この抽出すべき要約セグメントの決定処理は、まず、要約セグメントの候補（以下、要約セグメント候補という）をリスト化し、このリスト化された要約セグメント候補から抽出すべき要約セグメントを絞り込んで決定するようになっている。

【0182】さらに、本実施形態では、入力されるデジタル音声映像情報には映像情報と付帯されている音声情報が多重化されているものとして、以下に詳細に説明する。

【0183】図1に示すように、本実施形態の要約再生装置100は、取得されたデジタル音声映像情報から音声情報およびジャンル情報を有する情報（以下、付帯情報という）を分離するデマルチプレクサ101と、分離された音声情報を復号して予め設定されたパラメータに基づいて音声特徴量を取得する音声特徴量抽出部102と、付帯情報からジャンル情報を取得するジャンル情報取得部103と、音声映像情報を蓄積するとともに、当該音声映像情報において取得された音声特徴量およびジャンル情報を蓄積する蓄積部104と、各部の操作を行うとともに、音声映像情報を要約する時間長を入力する操作部105と、ジャンル情報に基づいて音声特徴量を抽出する際の最適な決定パラメータを設定する決定パラメータ設定部106と、蓄積された音声映像情報の要約再生を行う再生部107と、決定パラメータに適合するパラメータによって抽出された音声特徴量に基づいて蓄積された音声映像情報における要約セグメントを決定するとともに再生部107を制御する制御部108と、要約再生された映像信号を音声信号とともに表示・出力する表示部109とを備えている。

【0184】なお、ジャンル情報取得部103および決定パラメータ設定部106は、それぞれ、本発明に係る取得手段および最適化手段を構成する。また、制御部108は、再生部107とともに本発明に係る生成手段を構成し、さらに、制御手段108は本発明に係る決定手段および設定手段を構成する。

【0185】デマルチプレクサ101には、通信回線か

ら送出された若しくは図示しない受信部によって受信されたデジタル音声映像情報、または、既に蓄積部104に蓄積されたデジタル音声映像情報が入力されるようになっており、このデマルチプレクサ101は、この入力されたデジタル音声映像情報から音声情報および付帯情報を見離して、それぞれ、音声特徴量抽出部102およびジャンル情報取得部103に出力するようになっている。

【0186】この付帯情報は、例えば、ニュース番組、スポーツ観戦番組、映画、ドラマ、歌番組およびバラエティ番組などのジャンル情報を一部に有しており、B S (Broadcasting Satellite) デジタル放送では、番組配列情報（Service Information）がこの付帯情報に当たる。この場合、デマルチプレクサ101は番組配列情報をジャンル情報取得部103に出力するようになっている。

【0187】音声特徴量抽出部102にはデマルチプレクサ101から出力されたデジタル信号である音声情報が入力されるようになっており、この音声特徴量抽出部102は、この入力された音声情報を復号するとともに、予め設定された複数のパラメータに基づいて当該音声情報の音声特徴量、例えば、音声レベルまたは周波数特性を取得するようになっており、この取得した音声特徴量を蓄積部104に出力するようになっている。

【0188】具体的には、音声特徴量抽出部102は、入力された音声情報に基づいて単位時間平均音圧レベル（パワー）を算出し、または、各周波数毎の音圧レベルを算出し、予め定められた複数のパラメータに基づいて複数の音声区間を抽出するようになっている。

【0189】本実施形態では、予め定められた複数の音声レベルをパラメータとして複数の音声区間を抽出するようになっており、例えば、音声の無い区間と閾値（音声レベル）が異なる複数の騒音区間などの音声区間を検出するようになっている。この音声区間の検出処理については後述する。

【0190】ジャンル情報取得部103にはデマルチプレクサ101から出力された付帯情報（番組配列情報）が入力されるようになっており、このジャンル情報取得部103は、この入力された付帯情報（番組配列情報）に基づいて音声映像情報における映画、ニュースなどジャンル情報を取得し、このジャンル情報を蓄積部104に出力するようになっている。

【0191】なお、このジャンル情報は、入力されたデジタル音声映像情報と別に取得するようにしてもよい。例えば、インターネットなどの通信回線を介して任意の場所に格納されている当該入力された音声映像情報に関するジャンル情報を取得するようにしてもよい。

【0192】また、ジャンル情報を入力された音声映像情報から取得する代わりに、当該入力された音声映像情報における音声情報に基づいて推定するようにしてもよ

い。

【0193】この場合、本実施形態に推定部を設け、この推定部によって入力された音声映像情報からジャンルを推定する。例えば、各ジャンルを特定するような特徴のあるテキスト情報を予め用意しておき、入力された音声映像情報から音声認識技術を用いて音声を抽出するとともに、その音声認識結果をテキスト情報として予め用意されたテキスト情報との一致度を算出し、この算出結果に基づいてジャンルを推定するようになっている。

【0194】蓄積部104には入力されたデジタル音声映像情報、音声特徴量102によって抽出された音声情報の音声特徴量およびジャンル情報取得部103によって取得された音声映像情報のジャンル情報を蓄積するようになっている。また、この蓄積部104は制御部108の指示に基づいて再生部107に音声映像情報の出力並びに音声特徴量およびジャンル情報の決定パラメータ設定部106および制御部108に出力するようになっている。

【0195】操作部105は、ユーザによって音声映像情報における蓄積制御指示、蓄積された音声映像情報の再生指示および要約再生を行う際の要約再生時間の入力などを行うことができるようになっており、この指示が制御部108に入力されると、この指示に基づいて制御部108が各部を制御するようになっている。

【0196】決定パラメータ設定部106には蓄積部104から出力されたジャンル情報が入力されるようになっており、この決定パラメータ設定部106は入力されたジャンル情報に基づいて制御部108によって抽出すべき要約セグメントの決定処理の際に使用する決定パラメータ、具体的には最適な音声レベルを設定し、この設定された決定パラメータの値（音声レベル）を制御部108に出力するようになっている。

【0197】また、決定パラメータ設定部106は、入力されたジャンル情報に基づいて抽出すべき要約セグメントを決定するときの重要度を最適化し、当該最適化した値または最適化のためのパラメータを制御部108に出力するようになっている。この重要度の最適化については後述する。

【0198】再生部107には蓄積部104から出力されたデジタル音声映像情報が入力されるようになっており、再生部107は入力された多重化されている音声映像情報を映像情報および音声情報に分離するとともに復号し、制御部108の指示に基づいて要約再生を行うようになっている。また、この再生部107はこの再生された音声信号および映像信号を表示部109に出力するようになっている。

【0199】制御部108は、操作部105によって入力された指示に基づいて蓄積部104における蓄積制御並びに音声特徴量抽出部102および決定パラメータ設定部106によって設定された音声特徴量およびパラメ

ータに基づいて後述する要約セグメントを決定するとともに、この決定された要約セグメントに基づいて再生部107の再生制御を行うようになっている。

【0200】表示部109には再生部107から出力された音声信号および映像信号が入力されるようになっており、この表示部109は、この入力された映像信号をモニター画面などに表示するとともに、音声信号をスピーカなどによって拡声するようになっている。

【0201】次に、図2、図3を用いて本実施形態の音声特徴量の抽出処理について説明する。

【0202】なお、図2および図3は本実施形態における無音区間および騒音区間の検出原理を説明するための図である。

【0203】通常、録画された音声映像情報、または、通信回線などを介して提供される音声映像情報において、その音声映像情報の時間長を短縮して要約を行う際に、当該音声映像情報に付帯される音声情報は重要な役割を果たしている。

【0204】例えば、テレビジョン放送番組において、騒音区間は番組の盛り上がった部分を、また、無音区間は場面が転換されたまたは内容が切り替わった部分を示す。

【0205】具体的には、ニュース番組であれば、ニュース内容が切り替わるときには無音部分、いわゆる「間（ポーズ）」が取られるとともに、その「間」に続く部分には後に続く内容が示され、その部分は映像情報の特徴部分となる。特に、無音区間に続く部分は次の内容の冒頭部分を示すので、当該内容の概略が示されることも多い。

【0206】また、スポーツ観戦番組であれば、歓声などの背景雑音には観客の反応が現れるので、盛り上がった場面（野球のホームランのシーンやサッカーのゴールシーン）では、自然にアナウンサーや観衆の声が大きくなり、音声レベルが他の部分のレベルに比べて非常に高くなり、その部分は映像情報の特徴部分となる。

【0207】したがって、ニュース番組などの無音区間を有する音声映像情報では無音区間の検出が重要となる。また、スポーツ観戦番組などの背景雑音に歓声を有する音声映像情報では、殆ど無音区間が検出されないとともに、閾値の異なる騒音区間を検出する方がより適した要約再生を行うことができるようになっている。

【0208】このように、異なるジャンルによって抽出すべき無音区間および騒音区間などの音声区間および当該音声区間における最適な閾値が異なる。

【0209】以上のことから、本実施形態では、抽出された音声情報における単位時間平均音圧レベル（パワー）を予め音声特徴量抽出部102において複数の閾値、例えば、音声レベルによって無音区間や騒音区間など複数の音声区間を抽出しておき、決定パラメータ設定部106において、入力されたジャンル情報に基づいて

39

この抽出するときの決定パラメータを設定して、制御部108において、当該設定された決定パラメータに適合するパラメータによって抽出された音声区間から、要約セグメントを決定する際に使用する最適な音声区間を選択するようになっている。

【0210】具体的には、決定パラメータ設定部106において、音声映像情報がスポーツ観戦番組である場合に、ニュース番組に比べ閾値（音声レベル）が高い騒音区間を選択するとともに、逆に、常にどの場面でも歓声音があるため、無音区間は殆ど検出されないので、スポーツ観戦番組では無音区間の処理を行わないように設定するようになっている。また、ニュース番組であれば、無音区間を検出する音声レベルの低い区間と標準的な騒音区間を採用するようになっている。

【0211】本実施形態では、無音区間と騒音区間を検出する場合は、音声特微量抽出部102は図2に示すように、予め設定された無音レベルの閾値（以下、無音レベル閾値（THs）という）および騒音レベルの閾値（以下、騒音レベル閾値（THn）という）に基づいて音声映像情報中における各無音区間および騒音区間の音声映像情報の時間軸上の開始位置（以下、単に開始位置という）および終了位置（以下、単に終了位置という）を検出するようになっている。本実施形態では、この各無音区間および騒音区間毎に検出した開始位置および終了位置の時間情報を蓄積部104に出力するようになっている。なお、各無音区間および騒音区間の時間的長さを以下区間長という。

【0212】具体的には、上述したように、入力された音声情報に基づいて単位時間平均音圧レベル（パワー）を算出し、この算出した値によって得られた音声情報が無音レベル閾値（THs）以下または騒音レベル閾値（THn）以上で、かつ、予め設定された時間的長さ（以下、最短無音区間長（DRSmin）および最短騒音区間長（DRNmin）という）以上の区間が検出された場合に、その区間を無音区間として、または、騒音区間として検出するようになっている。

【0213】なお、この音声レベルの標準的な値として、ニュース番組ではアナウンサーが声を発したときの音声は、-50dB以上になるので、無音レベル閾値（THs）を-50dBと設定し、スポーツ番組では観衆が盛り上がった際には背景雑音の音声レベルが-35dB程度になるので、騒音レベル閾値（THn）を-35dBと設定するようになっており、複数の無音区間および騒音区間を検出するときはこの音声レベルを基準として複数設定するようになっている。

【0214】また、本実施形態では、ニュース番組およ\*

$$\text{重要度 } f(x) = ax + b \quad \dots \quad (\text{式1})$$

$$\text{重要度 } f(x) = x + 100 \quad \dots \quad (\text{式2})$$

なお、xは、通常、設定される重要度の値であり、本実施形態では各音声区間の区間長を使用するようになって

40

\*びスポーツ観戦番組の音声レベルを考慮して、最短無音区間長（DRSmin）および最短騒音区間長（DRNmin）をそれぞれ0.2秒、1.0秒に設定している。

【0215】さらに、図3は、複数の騒音レベル閾値（THn）によって複数の騒音区間を検出する場合の図であり、複数の騒音区間を検出する場合は、図3に示すように、異なる音声レベル、すなわち、騒音レベル閾値1（THn1）および騒音レベル閾値2（THn2）によって騒音区間1および騒音区間2を検出するようになっている。

【0216】次に、本実施形態の重要度の最適化について説明する。

【0217】上述のように、テレビジョン放送番組において、騒音区間は番組の盛り上がった部分を、また、無音区間は場面が転換されたまたは内容が切り替わった部分を示す一方、スポーツ観戦番組であれば、歓声などの背景雑音には観客の反応が現れるので、盛り上がった場面では、音声レベルが他の部分のレベルに比べて非常に高くなり、その部分は映像情報の特徴部分となる。

【0218】また、ニュース番組であれば、ニュース内容が切り替わるときには、無音部分、いわゆる「間（ポーズ）」が取られるとともに、その「間」に続く部分には後に続く内容が示され、その部分は映像情報の特徴部分となる。特に、無音区間に続く部分は次の内容の冒頭部分を示すので、当該内容の概略が示されることも多い。

【0219】すなわち、ニュース番組などのように無音区間を有する音声映像情報では、無音区間の方が騒音区間に比べて重要度が高くなる。また、ニュース番組のように完成の背景雑音を有する音声映像情報では、騒音区間の方が無音区間に比べて重要度が高くなる。

【0220】このようにジャンル毎に音声区間によって設定される要約セグメントの重要度が異なるので、本実施形態では、決定パラメータ設定部106によって取得したジャンル情報に基づいてジャンル毎に重要度を最適化するようになっている。

【0221】具体的には、ジャンル情報取得部103によってニュース番組などの無音区間を有する音声映像情報のジャンルであることが取得された場合には、無音区間によって決定された要約セグメントの重要度を以下のような重み付け関数によって設定するようになっており、本実施形態では、このパラメータを制御部108に出力するようになっている。

【0222】

$$\dots \quad (\text{式1})$$

$$\dots \quad (\text{式2})$$

いる。また、a、bは定数である。

50 【0223】また、スポーツ観戦番組などの背景雑音に

41

歓聲音を有する音声映像情報である場合は、複数の騒音区間、例えば、図3に示す騒音区間1および騒音区間2を選択するとともに、閾値の高い騒音区間1によって設定された要約セグメントの重要度を、無音区間に基づいて決定された要約セグメントの重要度を設定するときに用いられた重み付け関数などによって騒音区間2によって決定される要約セグメントの重要度より高く設定するようになっている。

【0224】次に、図4、図5を用いて本実施形態における抽出すべき要約セグメントの決定処理について説明する。

【0225】なお、図4は騒音区間に基づいてセグメントの開始時刻および終了時刻を決定する原理を説明するための図であり、図5は無音区間に基づいてセグメントの開始時刻および終了時刻を決定する原理を説明するための図である。

【0226】上述のように、ニュース番組において、ニュース内容が切り替わるとには、無音部分、いわゆる「間（ポーズ）」が取られるとともに、その「間」に続く部分には後に続く内容が示され、その部分は映像情報の特徴部分となるため、無音区間に続く部分が重要になる。

【0227】また、スポーツ観戦番組であれば、歓声などの背景雑音には観客の反応が現れるので、盛り上がった場面では、音声レベルが他の部分のレベルに比べて非常に高くなり、その部分は映像情報の特徴部分となる。

【0228】このように、無音区間または騒音区間と音声映像情報の特徴部分との時間軸上の位置関係および重要度が異なるので、抽出すべき要約セグメントを決定する際に無音区間と騒音区間によって決定する処理を別の処理によって行うようになっている。以下に、本実施形態の要約セグメントの決定処理について説明する。

【0229】なお、本実施形態における要約セグメントの決定処理では、無音区間および騒音区間に基づいて要約セグメントの開始時刻（ $S_{TSSi}$ ）、終了時刻（ $S_{ESSi}$ ）および重要度（ $IPSSi$ ）を決定するようになっている。ただし、 $i$ は $i$ 番目の無音区間および騒音区間であり、 $j$ は $j$ 番目の要約セグメントであることを示す。

【0230】また、本実施形態における要約セグメントの決定処理では、まず、無音区間および騒音区間に基づいて要約セグメントの開始時間および重要度を決定して要約セグメント候補をリスト化し、その後、要約セグメント候補の絞り込みを行い、最短要約セグメント時間長、標準要約セグメント時間長および最長要約セグメント時間長を決定して絞り込まれた要約セグメント候補の終了時刻を決定するようになっている。

【0231】さらに、本実施形態の要約セグメントの決定処理では、要約セグメントの決定する際に基になった無音区間および騒音区間の区間長情報（ $DRSSi$ ）を

42

保持するようになっている。本実施形態では、要約セグメントが一度決定され、要約セグメントの絞り込み処理を行った後、終了時刻を決定するため、この区間長情報（ $DRSSi$ ）は、後述する終了時刻を決定する際に、無音区間に基づいて決定された要約セグメントであるか、騒音区間にによって決定された要約セグメントであるがを判断する必要があり、その判断にこの区間長情報（ $DRSSi$ ）を用いるようになっている。

【0232】具体的には、本実施形態では、騒音区間に基づいて設定された要約セグメントには、基準となった騒音区間の区間長を設定するようになっている（ $DRDNi = DRSSi$ ）。また、無音区間に基づいて設定された要約セグメントには、 $DRSSi = 0$ と設定するようになっている。

【0233】したがって、要約セグメント決定処理の動作において、後述する終了時刻を決定する際に、この $DRSSi$ が「0」であれば、無音区間に基づいて、「 $DRSSi \neq 0$ 」であれば、騒音区間に基づいて決定された要約セグメントと判断できるようになっている。

【0234】【騒音区間における要約セグメントの設定】上述のように、騒音区間は番組の盛り上がった部分を示すので、騒音区間が重要である。そこで、本実施形態では、図4に示すように、検出部103によって検出された騒音区間の開始位置を要約セグメントの開始位置と設定するようになっている。

【0235】なお、スポーツ観戦番組では、観客の歓声が集音され、この集音された音が背景雑音として当該音声映像情報に付帯されている音声情報に含まれている場合には、場面が盛り上がる少し前から再生する方が要約再生する際には有効性が高い。通常、スポーツ観戦において好プレーや得点シーンなど盛り上がる部分では観客が歓声を上げるまで、すなわち、騒音区間となるまでに時間差が生ずる。したがって、スポーツ観戦番組などの音声映像情報において騒音区間に基づいて要約セグメントの開始時刻を騒音区間の開始位置から $\Delta t$ だけ前にずらすようにしてもよい。

【0236】また、騒音区間における要約セグメントの終了時刻は、騒音区間の終了位置に基づいて決定されるようになっている。

【0237】抽出される要約セグメントの内容を考慮すると、原則的には騒音区間の終了位置を要約セグメントの終了時刻に設定する必要がある。しかしながら、抽出すべき要約セグメントは、あまりにも要約セグメントの時間長が短いと当該場面を理解することが困難になる一方、不用意に長い時間長を有する場合には、無駄な情報が多くなるとともに、情報量が増大し映像情報を的確に要約することができない。

【0238】このため、後述する最短要約セグメント時間長（ $DR_{Min}$ ）、標準要約セグメント時間長（ $DR_{Typ}$ ）および最長要約セグメント時間長（ $DR_{Max}$ ）を

43

設定し、これらの時間長を用いて要約セグメントの終了時刻を設定するようになっている。

【0239】例えば、図4に示すように、騒音区間(DNi)（例えば、図4に示す騒音区間a）が最短要約セグメント時間長(DRMin)に満たない場合は、最短要約セグメント時間長(DRMin)が要約セグメントの時間長になり、要約セグメントの開始時刻にその最短要約セグメント時間長(DRMin)を加えた時刻を要約セグメントの終了時刻に設定するようになっている。

【0240】また、騒音区間(DNi)（例えば、図4に示す騒音区間b）が最短要約セグメント時間長(DRMin)以上、最長要約セグメント時間長(DRMax)以下\*

$$0 < DRSS_i < DR_{Min} \text{ の場合, } SESSi = STSS + DR_{Min} \quad \dots \quad (\text{式3})$$

$$DR_{Min} \leq DRSS_i \leq DR_{Max} \text{ の場合, } SESSi = STSS + DRSS_i \quad \dots \quad (\text{式4})$$

$$DR_{Max} < DRSS_i \text{ の場合, } SESSi = STSS + DR_{Typ} \quad \dots \quad (\text{式5})$$

となる。ただし、上述のように要約セグメントの開始時刻を $\Delta t$ 早めたときは、他の要約セグメントの時間長との整合性により、各最短要約セグメント時間長(DRMin)、標準要約セグメント時間長(DRTyp)および最長要約セグメント時間長(DRMax)の時間長から $\Delta t$ 分減算した時間を各最短要約セグメント時間長(DRMin)、標準要約セグメント時間長(DRTyp)および最長要約セグメント時間長(DRMax)にするようになっている。

【0243】なお、本実施形態において、各要約セグメントの終了時刻の設定は、後述する要約セグメント候補の絞り込み処理が行われ、当該絞り込まれた要約セグメントに対して設定されるようになっている。すなわち、騒音区間に基づいて要約セグメントの開始時刻を設定することによって当該要約セグメントを要約セグメント候補としてリストアップし、その後、後述する要約セグメント候補の絞り込み処理が行われ、さらに、最短要約セグメント時間長(DRMin)、標準要約セグメント時間長(DRTyp)および最長要約セグメント時間長(DRMax)が設定された後に当該要約セグメントの終了時刻を設定するようになっている。

【0244】また、騒音区間における要約セグメントの重要度(IPSSi)は、基本的には騒音区間の区間長(DRDNi)を用いて設定するようになっており、騒音区間の区間長が長ければ長いほど重要度を高く設定することができるようになっているが、決定パラメータ設定部106によって各要約セグメントに設定する重要度が最適化された場合にはその指示に従うようになっている。

【0245】〔無音区間における要約セグメントの設定〕上述のように無音区間は、場面が転換されたまたは内容が切り替わった部分を示すので、無音区間が終了し

44

\*である場合は、騒音区間長が要約セグメントの時間長となり、騒音区間が終了する位置を要約セグメントの終了時刻に設定するようになっている。

【0241】さらに、騒音区間(DNi)（例えば、図4に示す騒音区間c）が最長要約セグメント時間長(DRMax)を越える場合は、要約セグメントの開始時刻に標準要約セグメント時間長(DRTyp)を加えた時刻を要約セグメントの終了時刻と設定するようになっている。

10 【0242】すなわち、i番目の騒音区間におけるj番目の要約セグメントにおいて、セグメント時間長(DRDNi=DRSSi)より、

$$DR_{Min} \leq DRSS_i \leq DR_{Max} \text{ の場合, } SESSi = STSS + DRSS_i \quad \dots \quad (\text{式4})$$

$$DR_{Max} < DRSS_i \text{ の場合, } SESSi = STSS + DR_{Typ} \quad \dots \quad (\text{式5})$$

た後に続く部分が重要である。そこで、本実施形態では、図5に示すように、検出部103によって検出された無音区間の予め設定された区間長（以下、追加最短無音区間長(DRSAmin)）以上、例えば、1.0秒以上を有する無音区間の終了位置を要約セグメントの開始時間(STSS)に設定するようになっている。

【0246】なお、無音区間には重要でないものも存在するため、明らかに「間」が存在し、内容が切り替わったところを検出するため、要約セグメントの決定にあたっては、追加最短無音区間長(DRSAmin)を定め、この追加最短無音区間長(DRSAmin)以上の区間長を有している無音区間の終了位置を要約セグメントの開始位置に設定するようになっている。

【0247】また、無音区間における要約セグメントの終了時刻は、要約セグメントの開始時刻の設定に用いられた無音区間の次の無音区間の開始位置に基づいて決定されるようになっている。

【0248】この場合、要約セグメントの開始時刻の設定に用いられた無音区間の次の無音区間の区間長は、追加最短無音区間長(DRSAmin)以上の長さである必要はなく、抽出された全ての無音区間を対象に無音区間の検索を行うようになっている。

【0249】騒音区間の場合と同様に要約セグメントの終了時刻は、後述する最短要約セグメント時間長(DRMin)、標準要約セグメント時間長(DRTyp)および最長要約セグメント時間長(DRMax)を用いて設定するようになっている。

【0250】例えば、図5に示すように、要約セグメントの開始時刻に設定された無音区間(DSi)の次に検出された無音区間(DSi+1)（例えば、図5に示す無音区間a）の開始位置が最短要約セグメント時間長(DRMin)に満たない場合には、最短要約セグメント時間

45

長 ( $D R_{Min}$ ) が要約セグメントの時間長になり、要約セグメントの開始時刻にその最短要約セグメント時間長 ( $D R_{Min}$ ) を加えた時刻を要約セグメントの終了時刻に設定するようになっている。

【0251】また、無音区間 ( $D S_i$ ) の次に検出された無音区間 ( $D S_{i+1}$  (例えば、図5に示す無音区間 b)) の開始位置が最短要約セグメント時間長 ( $D R_{Min}$ ) を越えるとともに、最長要約セグメント時間長 ( $D R_{Max}$ ) に満たない場合は、検出された無音区間 ( $D S_{i+1}$ ) の開始位置を要約セグメントの終了時刻と設定するようになっている。

【0252】さらに、無音区間 ( $D S_i$ ) の次に検出された無音区間 ( $D S_{i+1}$  (例えば、図5に示す無音区間 c)) の開始位置が最長要約セグメント時間長 ( $D R_{Max}$ ) を越える場合は、標準要約セグメント時間長 ( $D R_{Typ}$ ) が要約セグメントの時間長になり、要約セグメントの開始時刻に標準要約セグメント時間長 ( $D R_{Typ}$ ) を加えた時刻を要約セグメントの終了時刻と設定するようになっている。

【0253】また、本実施形態では、最短要約セグメント時間長 ( $D R_{Min}$ ) 、標準要約セグメント時間長 ( $D R_{Typ}$ ) を加えた時刻を要約セグメントの終了時刻と設定するようになっている。

$$S E S S_j = S T$$

また、区間 [ $D R_{Min}$ 、 $D R_{Max}$ ] に無音区間 ( $D S_{i+1}$ ) の開始位置 ( $S T$ ) が見つからず、区間 [0、 $\ast$

$$S E S S_j = S T S S_i$$

さらに、区間 [0、 $D R_{Max}$ ] に無音区間 ( $D S_{i+1}$ ) の

$$S E S S_j = S T S S_i$$

となる。

【0256】この無音区間 ( $D S_{i+1}$ ) の検出の順序は、最短要約セグメント時間長 ( $D R_{Min}$ ) 以内に次の無音区間 ( $D S_{i+n}$ ) が存在しても、無音区間 (例えば、無音区間 ( $D S_{i+n}$  ( $n \geq 2$ ))) の開始位置が最短要約セグメント時間長 ( $D R_{Min}$ ) 以上、最長要約セグメント時間長 ( $D R_{Max}$ ) 以下に存在した場合には、最短要約セグメント時間長 ( $D R_{Min}$ ) 以内に存在する次の無音区間 ( $D S_{i+1}$ ) を要約セグメントの開始時刻の基準となった無音区間 ( $D S_i$ ) の次の無音区間として取り扱わず、無音区間 ( $D S_{i+n}$  ( $n \geq 2$ )) を無音区間 ( $D S_{i+1}$ ) として当該無音区間 ( $D S_{i+1}$ ) に基づいて要約セグメントの終了時刻を決定するようになっている。

40

★

$$I P S S_j = f(D R D S_i)$$

なお、 $f(\cdot)$  は重み付け関数であり、本実施形態では、上述のように、(式1) または(式2) を用いるようになっている。

【0260】〔要約セグメント候補の絞り込み処理〕上述のように無音区間および騒音区間に基づいて決定された要約セグメント全てについて後述する要約再生の処理を行ってもよいが、処理量の軽減および不必要的要約セグメントによる要約再生を防ぐ、すなわち、重要度の低

46

\*  $R_{Typ}$  および最長要約セグメント時間長 ( $D R_{Max}$ ) を用いて要約セグメントの終了時刻を設定する際に、当該次の無音区間の検出において以下の順序によって行うようになっている。

【0254】要約セグメントの開始時間の基準となった無音区間 ( $D S_i$ ) の次にある無音区間 ( $D S_{i+1}$ ) を検出する順序としては、まず、無音区間 ( $D S_i$ ) の次に検出された無音区間 ( $D S_{i+1}$ ) の開始位置が最短要約セグメント時間長 ( $D R_{Min}$ ) 以上、最長要約セグメン

10 ト時間長 ( $D R_{Max}$ ) 以下に存在するか否かを検出し、この範囲内に無いとき、無音区間 ( $D S_i$ ) の次に検出された無音区間 ( $D S_{i+1}$ ) の開始位置が最短要約セグメント時間長 ( $D R_{Min}$ ) までに無いか否かを検出するようになっている。さらに、これらの範囲に無かったときに、無音区間 ( $D S_i$ ) の次に検出された無音区間 ( $D S_{i+1}$ ) は最長要約セグメント時間長 ( $D R_{Max}$ ) 以上にあると判断するようになっている。

【0255】すなわち、 $i$  番目の騒音区間における  $j$  番目の要約セグメントにおいて、区間 [ $D R_{Min}$ 、 $D R_{Max}$ ] に無音区間 ( $D S_{i+1}$ ) の開始位置 ( $S T$ ) が見つかった場合、

... (式6)

\*  $D R_{Min}$  に見つかった場合、

+  $D R_{Min}$  ... (式7)

開始位置 ( $S T$ ) が見つから無かった場合、

+  $D R_{Typ}$  ... (式8)

★【0257】なお、騒音区間における要約セグメントの終了時刻の設定と同様に、無音区間によって各要約セグメントの終了時刻の設定は、後述する要約セグメント候補の絞り込み処理が行われ、当該絞り込まれた要約セグメントに対して設定されるようになっている。

【0258】また、無音区間における重要度は、基本的には騒音区間と同様に無音区間の区間長 ( $I P S S_j$ ) に基づいて設定するが、決定パラメータ設定部106によって各要約セグメントに設定する重要度が最適化された場合にはその指示に従うようになっており、例えば、以下のよう (式9) によって算出するようになっている。

【0259】

... (式9)

い要約セグメントであっても、後述する結合処理によって重要度が高くなる可能性があり、不適切な要約セグメントによる要約再生を防ぐため、決定された要約セグメントを絞り込むようになっている。

【0261】本実施形態では、以下の (式10) によってリスト化された要約セグメント候補から絞り込み処理を行うようになっている。

50 【0262】この (式10) は、全ての要約セグメント

47  
の時間が限界最短時間（ $DR_{Min}$ ）とした場合の絞り込む要約セグメントの数の定数倍（例えば、 $K_1 = 2$ ）と要約セグメント候補の数とを比較し、少ない方を要約セグメント数に設定するようになっている。  
\* \*

$$NP_{new} =$$

$$\min(\text{Int}(k_1 \times (S / DR_{Min})), NP_{old}) \dots \text{ (式10)}$$

となる。

【0264】なお、 $k_1$ は定数であり、 $\min(a, b)$ は $a, b$ のうち値の小さい方を選択することを示し、 $\text{INT}(\cdot)$ は小数点以下の切り捨てを行うことを示す。また、 $NP_{new}$ は絞り込まれた数を示し、 $DR_{Min}$ は限界最短時間を示す。

【0265】この限界最短時間（ $DR_{Min}$ ）とは、人間が一つの要約セグメント内容を理解するのに必要最低限の時間であり、例えば、本実施形態では限界最短時間（ $DR_{Min}$ ）は4秒としている。

【0266】このように算出された要約セグメント候補の数が絞り込む要約セグメントの数の定数倍より多い場合は、すなわち、 $NP_{new} < NP_{old}$ の場合は、重要度順に要約セグメント候補を $NP_{new}$ 個選択し、それ以外は、要約セグメント候補から削除するようになっている。

【0267】なお、本実施形態では、このように要約セグメント候補の絞り込みを行い、この絞り込まれた要約セグメント候補において、上述した設定方法によって各要約セグメントに終了時刻を設定するようになっている。

【0268】【最短／標準／最長要約セグメント時間長の設定】上述したように、抽出すべき要約セグメント

48  
\* 【0263】例えば、リスト化された要約セグメント候補の数（ $NP_{old}$ ）とし、要約時間 $S$ 、とすると、新たに設定される要約セグメント候補の数（ $NP_{new}$ ）は、

$$NP_{new} =$$

$$\min(\text{Int}(k_1 \times (S / DR_{Min})), NP_{old}) \dots \text{ (式10)}$$

※は、なるべく長い時間長を有する方が当該場面を理解する上で必要である一方、不用意に長い時間長を有する場合には、無駄な情報が多くなるとともに、情報量が増大し映像情報を的確に要約することができないので、本実施形態では、後述する最短要約セグメント時間長（ $DR_{Min}$ ）、標準要約セグメント時間長（ $DR_{Typ}$ ）および最長要約セグメント時間長（ $DR_{Max}$ ）を設定するようになっている。

【0269】例えば、本実施形態では、抽出すべき要約セグメントの各内容を各内容毎に的確に把握させるため、以下の式によって最短要約セグメント時間長（ $DR_{Min}$ ）、標準要約セグメント時間長（ $DR_{Typ}$ ）および最長要約セグメント時間長（ $DR_{Max}$ ）を設定するようになっている。

【0270】最短要約セグメント時間長（ $DR_{Min}$ ）についてはユーザにおける要約セグメントの見やすさを考慮して、比較的要約セグメントの時間長を長くするため（式11）に示すようになっており、また、標準要約セグメント時間長（ $DR_{Typ}$ ）および最長要約セグメント時間長（ $DR_{Max}$ ）については、（式11）によって算出された最短要約セグメント時間長（ $DR_{Min}$ ）を定数倍することによって算出するようになっている。

$$DR_{Min} = \max(DR_{Min}, (K_2 \times (S / NP_{new}))) \dots \text{ (式11)}$$

$$DR_{Typ} = DR_{Min} \times K_1$$

$$DR_{Max} = DR_{Min} \times K_2$$

【0271】なお、 $K_1$ および $K_2$ とも比例定数であり、 $\max(a, b)$ は $a, b$ のうち値の大きい方を選択することを示す。また、 $K_2 (\geq 1)$ は各要約セグメントの最短時間を決定する係数であり、この値が大きいほど最短時間が長くなるとともに、要約セグメントの数は減少するようになっている。例えば、本実施形態では、 $K_2 = 1, 2, K_1 = 2, K_2 = 3$ と設定するようになっている。

★40

★【0272】【要約セグメントの結合】ここで、本実施形態では、時間的に重なる複数の要約セグメントが有った場合にこれらの要約セグメントを一つの要約セグメントに結合するようになっており、この場合、結合することによって生成された要約セグメントの重要度は、各要約セグメントが有する重要度（ $IPSS_j$ ）の高い方を選択するようになっている。

$$IPSS_j = \max(IPSS_j, IPSS_{j+n}) \dots \text{ (式14)}$$

また、例えば、2つの要約セグメント $SS_j, SS_{j+n}$ （ $STSS_j < STSS_{j+n}$ ） $SESS_j \geq STSS \star$   
 $SESS_j = SESS_{j+n}$

となる。

【0273】このため、たとえ要約セグメントの重要度が低い場合であっても、重要度の高い要約セグメントと時間的に重なっている場合に、重要度の高い要約セグメントを補完できるようになっている。

【0274】【要約セグメントの決定】本実施形態で

48

は、最後に指定された要約時間を達成するために、重要度順に要約セグメント候補の選択を行なっている。また、選択された要約セグメント候補の総時間が指定された要約時間を超えるまで選択し、この選択されたセグメントを要約セグメントとするようになっている。

50 【0275】ここで、重要度順に要約セグメントを決定

49

する場合に、各要約セグメントは、異なるセグメント時間長を有するので、指定された要約時間を越えることがある。この要約時間が越えることが問題となるときは、例えば、オーバーした時間を決定された要約セグメント数で分配し、それぞれの要約セグメントの終了時刻から分配された時間長を削除する処理を行うようになっている。

【0276】次に、図6を用いて本実施形態の要約再生動作について説明する。

【0277】なお、図6は本実施形態における要約再生動作を示すフローチャートである。また、本動作は、要約再生に必要な音声映像情報が既に蓄積部104に蓄積されているものとし、ユーザが要約再生を指示したときに本動作を行うようになっている。

【0278】まず、操作部105によってユーザから要約再生の指示が入力されると、音声特微量抽出部102は蓄積部104からデマルチプレクサ101を介して音声映像情報における音声情報が入力され、予め設定されたパラメータに基づいて複数の音声区間、すなわち音声特微量を取得するとともに（ステップS11）、ジャンル情報取得部105は蓄積部104からデマルチプレクサ101を介して音声映像情報における番組配列情報からジャンル情報を取得する（取得処理工程（ステップS12））。

【0279】次いで、決定パラメータ設定部106は蓄積部104を介してジャンル情報を取得して最適な決定パラメータを設定して制御部108に出力するとともに、ジャンル情報に基づいて抽出すべき要約セグメントの重要度を設定する際のパラメータを設定する（最適化処理工程（ステップS13））。

【0280】次いで、制御部108は、決定パラメータ設定部106により設定された決定パラメータに適合するパラメータによって検出された音声区間を選択するとともに、当該選択された音声区間および決定パラメータ設定部106によって設定された重要度に基づいて上述の要約セグメントの決定処理を行う（決定処理工程（ステップS14））。

【0281】最後に、ステップ14によって抽出すべき要約セグメントが決定されると、制御部108は、再生部107を制御して、要約再生を開始するとともに、この決定された要約セグメントに基づいて再生部107を制御して要約再生を行う（生成処理工程（ステップS15））。

【0282】以上により本実施形態によれば、要約セグメントを決定するときの音声特微量、すなわち、基準となる複数の音声区間を検出するときの音声レベルおよび各要約セグメントの重要度を音声映像情報におけるジャンル情報に基づいて最適化することができるので、音声映像情報の種別が異なっても抽出すべき要約セグメントを的確に抽出することができ、音声映像情報の内容に即

50

した要約情報を再生することができる。

【0283】また、各閾値を最適化するのみで抽出すべき要約セグメントを的確に抽出することができるので、映像情報の種別毎に要約セグメントの決定処理処理動作の手順を変更すること無く容易に要約セグメントを決定することができる。

【0284】また、検出した複数の音声区間の時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、音声映像情報の盛り上がった部分および内容が切り替わった部分を要約セグメントとして的確に抽出することができ、音声映像情報の内容に即した要約再生を行うことができる。

【0285】また、異なる音声区間毎に最適化した重要度を各要約セグメントに設定することができるので、音声映像情報の種別が異なっても抽出すべき要約セグメントを的確に抽出することができ、音声映像情報の内容に即した要約再生を行うことができる。

【0286】なお、本実施形態はデジタル信号によって構成された映像情報に基づいて要約再生を行うようになっているが、テレビジョン放送によって提供された番組などの映像情報にも適用することを可能である。

【0287】また、本実施形態では、予め設定されたパラメータによって複数の音声区間を検出し、ジャンル情報に基づいて設定された決定パラメータに適合するパラメータによって検出された音声区間を選択して要約セグメントの決定処理を行うようになっていたが、上述のように、ジャンル情報を別に取得した場合には、取得したジャンル情報によって音声特微量抽出部102における音声区間を検出するパラメータを最適化するようにしてもよい。

【0288】また、本実施形態では、予め複数のパラメータによって複数の音声区間を抽出し、ジャンル情報に基づいて設定された決定パラメータに基づいて適合するパラメータによって抽出された音声区間にに基づいて要約セグメントの決定処理を行うようになっているが、デジタル音声映像情報と別に取得することによって予め当該音声映像情報のジャンル情報を取得することができる場合には、取得したジャンル情報を基づいて決定パラメータを設定し、当該決定パラメータに基づいて複数の音声区間を抽出するようにしてもよい。

【0289】また、本実施形態では、要約再生処理動作を、音声特微量抽出部102、ジャンル情報取得部103、決定パラメータ設定部106、再生部107および制御部108によって行うようになっているが、要約再生処理動作を行うプログラムを記録媒体に書き込むとともに、当該記録媒体に格納されたプログラムをコンピュータによって読み出すことによって要約再生を実行するようにしてもよい。

【0290】この場合、制御部108にこのプログラムを読み込み実行するコンピュータを備えるとともに、こ

51

のコンピュータによって音声特徴量の抽出、ジャンル情報の取得、パラメータの設定、要約セグメント決定処理および要約再生を行うようになる。この結果、上述した要約再生装置と同様の効果を得ることができる。

【0291】(II) 変形形態

次に、本発明に係る変形形態について説明する。

【0292】上述した実施形態においては、音声情報について、その騒音区間又は無音区間を用いて要約再生を行ったが、これ以外に、映像情報について、そのシーンチェンジ(場面変更)を利用して要約再生を行うことができる。

【0293】すなわち、具体的には、先ず、映像情報中におけるシーンチェンジを検出し、その後、当該検出した各シーンチェンジのタイミングを開始時刻として固定時間(例えば10秒)だけの再生を繰り返すことで上述してきた要約再生を行うように構成することが可能である。

【0294】ここで、当該要約再生に必要な全体の時間を調整するために、シーンチェンジの検出に対して重み付け(重要度の差を設けること)を行うように構成しても良い。そして、その重要度の決定には直前のシーンチェンジとの時間間隔を利用することが望ましく、当該重要度が高い順に要約再生で使用すべきシーンチェンジを決定するのである。更に、この重み付け関数をジャンル情報を用いて最適化するように構成することもできる。

【0295】次に、より具体的に二つの例について説明する。

【0296】[ニュース番組の場合] 初めに、ニュース番組の場合を例として説明する。

【0297】ニュース番組の内容を広く浅く拾い読みする(すなわち、要約再生する)には、なるべく多くのニュース内容を選択すべく、シーンチェンジ間隔が長いときのその後の部分を選んで再生することが望ましい。一方、逆に頻繁にシーンチェンジが為されるときはその後の部分は内容的にはほぼ同一であることが多い。よって、シーンチェンジの間隔が長い場合は重要度が高く、シーンチェンジの間隔が短い場合は重要度が低いと判断し、例えば  $f(x) = a \times x + b$  形式の演算式により重要度を決定するのがよい。

【0298】[スポーツ番組の場合] 次に、スポーツ番組の場合を例として説明する。

【0299】例えば、野球放送におけるピッチングのシーンや、サッカー放送におけるパスのシーン等、要約再生ではあまり重要でないシーンは、夫々のシーンチェンジの間隔が短い場合が多い。一方、野球放送におけるヒットのシーンや、サッカー放送におけるゴールのシーン等、要約再生で重要な部分は、夫々のシーンの巻き戻し再生や対象者のズームアップなどシーンチェンジが頻発する場合が多い。よって、シーンチェンジの間隔が長い場合はその重要度が低く、シーンチェンジの間隔が短い

52

場合はその重要度が高いと判断し、例えば  $f(x) = (a/x) + b$  形式の演算式により重要度を決定するのがよい。

【0300】また、本実施形態では、要約再生装置100が、上述のようにジャンル情報取得部103、決定パラメータ部106、再生部107および制御部108などによって構成されているが、制御部108にコンピュータおよびハードディスクなどの記録媒体を備え、ジャンル情報取得部103、決定パラメータ部106、再生部107および制御部108など要約再生装置100の各部に対応する上記処理を行うプログラムを当該記録媒体に格納し、当該プログラムをコンピュータに読み取らせることによって、上述したジャンル情報取得部103、決定パラメータ部106、再生部107および制御部108など要約再生装置100の各部のそれぞれの動作を行うようにもよい。この場合、上述した要約再生動作を行うときは、当該格納されたプログラムによりコンピュータを動作させることによって当該要約再生動作を行うようとする。また、この場合、制御部108は、本発明に係る取得手段、最適化手段、設定手段、生成手段および決定手段を構成する。

【0301】

【発明の効果】以上説明したように、請求項1記載の発明によれば、部分映像情報を決定するときに基準となる複数のコンテンツ区間を分類する際の閾値を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0302】また、請求項2記載の発明によれば、部分映像情報を決定するときに基準となる複数の音声区間を分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0303】また、請求項3に記載の発明によれば、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0304】また、請求項4に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取得することができるので、映像情報の内容に即した要約情報を得ることができる。

【0305】また、請求項5に記載の発明によれば、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歓声音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歓声音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができる。

【0306】また、請求項6に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0307】また、請求項7に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0308】また、請求項8に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0309】また、請求項9に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0310】また、請求項10に記載の発明によれば、映像情報の識別情報により、音声情報を構成する背景雑音に歓声音を有する映像情報であることが識別された場合に、歓声音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0311】また、請求項11に記載の発明によれば、部分映像情報を決定するときに基準となる複数のコンテンツ区間を分類する際の閾値を、コンテンツ情報における

識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0312】また、請求項12に記載の発明によれば、部分映像情報を決定するときに基準となる複数の音声区間を分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができ。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0313】また、請求項13に記載の発明によれば、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0314】また、請求項14に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取得することができるので、映像情報の内容に即した要約情報を得ることができる。

【0315】また、請求項15に記載の発明によれば、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歓声音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歓声音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができる。

【0316】また、請求項16に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0317】また、請求項17に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0318】また、請求項18に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0319】また、請求項19に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0320】また、請求項20に記載の発明によれば、映像情報の識別情報により、音声情報を構成する背景雑音に歓聲音を有する映像情報であることが識別された場合に、歓聲音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0321】また、請求項21に記載の発明によれば、部分映像情報を決定するときに基準となる複数のコンテンツ区間を分類する際の閾値を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、コンテンツ情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0322】また、請求項22に記載の発明によれば、部分映像情報を決定するときに基準となる複数の音声区間を分類する際の閾値を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。また、閾値を最適化するのみで抽出すべき部分映像情報を的確に抽出することができるので、映像情報の種別毎に部分映像情報の処理動作を変更すること無く容易に部分映像情報を決定することができる。

【0323】また、請求項23に記載の発明によれば、分類した複数の音声区間の少なくとも何れか一種類の映像情報中における少なくとも時間軸上の位置に基づいて抽出すべき部分映像情報を決定することができるので、映像情報の盛り上がった部分および内容が切り替わった部分を部分映像情報として的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0324】また、請求項24に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、この映像情報に適切な閾値によって的確に無音部分を有する無音区間を取得することができるので、映像情報の内容に即した要約情報を得ることができる。

【0325】また、請求項25に記載の発明によれば、映像情報の識別情報により、スポーツ観戦番組などの背景雑音に歓聲音を有する映像情報であることが識別された場合に、この映像情報に最適化された閾値によって的確に歓聲音区間を検出することができるので、映像情報の内容に即した要約情報を得ることができる。

【0326】また、請求項26に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、コンテンツ情報における識別情報に基づいて最適化することができるので、コンテンツ情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、コンテンツ情報の内容に即した要約情報を得ることができる。

【0327】また、請求項27に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0328】また、請求項28に記載の発明によれば、部分映像情報によって要約情報を生成する際の重要度を、映像情報における識別情報に基づいて最適化することができるとともに、異なる音声区間毎に最適化した重要度を各部分映像情報に設定することができるので、映像情報の種別が異なっても抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0329】また、請求項29に記載の発明によれば、映像情報の識別情報により、ニュース番組などの無音部分を有する映像情報であることが識別された場合に、無音部分を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

【0330】また、請求項30に記載の発明によれば、映像情報の識別情報により、音声情報を構成する背景雑音に歓聲音を有する映像情報であることが識別された場合に、歓聲音を有する区間によって決定された部分映像情報の重要度を最適化することができるので、抽出すべき部分映像情報を的確に抽出することができ、映像情報の内容に即した要約情報を得ることができる。

#### 【図面の簡単な説明】

【図1】本発明に係る要約再生装置の構成を示すブロック図である。

57

【図2】一実施形態における無音区間および騒音区間の検出原理を説明するための図である。

【図3】一実施形態において複数の騒音区間を検出する原理を説明する図である。

【図4】騒音区間に基づいてセグメントの開始時刻および終了時刻を決定する原理を説明する図である。

【図5】無音区間に基づいてセグメントの開始時刻および終了時刻を決定する原理を説明する図である。

【図6】一実施形態における要約再生動作を示すフローチャートである。

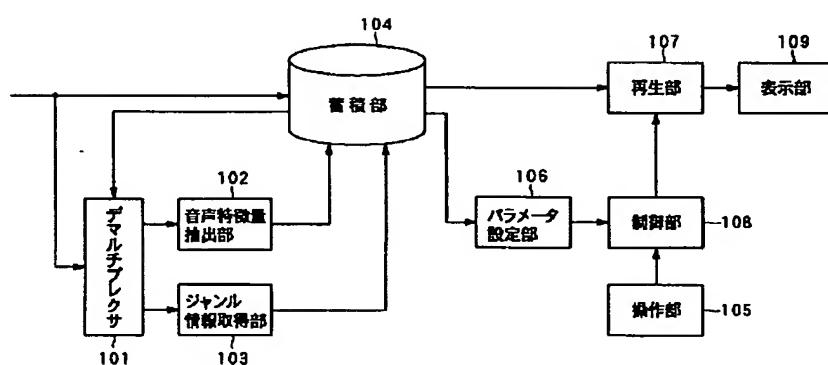
【図7】従来の要約再生装置の構成を示すブロック図である。

58

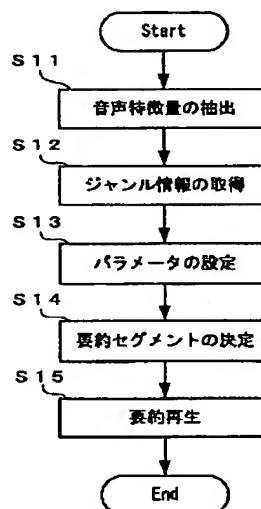
## 【符号の説明】

100 … 要約再生装置  
 101 … デマルチブレクサ  
 102 … 音声特微量抽出部  
 103 … ジャンル情報取得部（取得手段）  
 104 … 蓄積部  
 105 … 操作部  
 106 … パラメータ設定部  
 107 … 制御部  
 108 … 再生部（生成手段）  
 109 … 表示部  
 10 … 決定パラメータ設定部（最適化手段）  
 108 … 制御部（取得手段、最適化手段、設定手段、決定手段、生成手段）

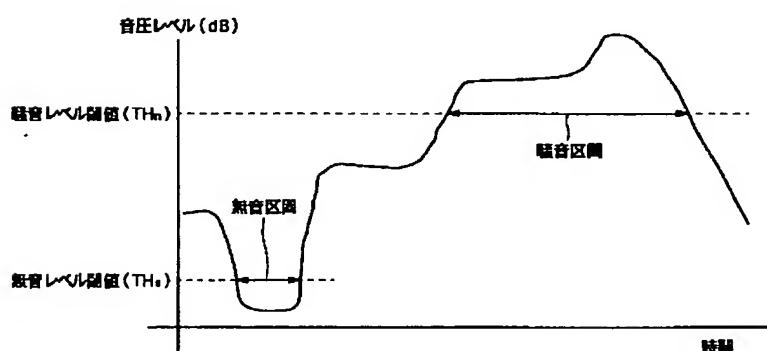
【図1】



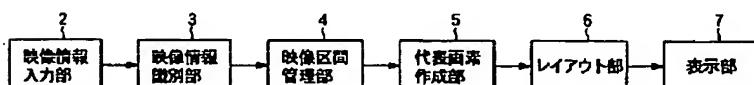
【図6】



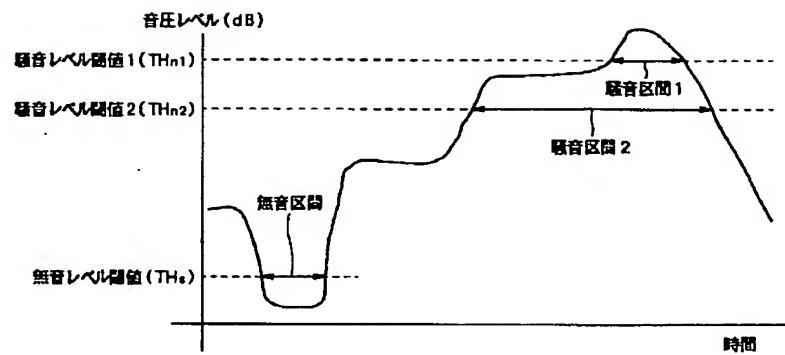
【図2】



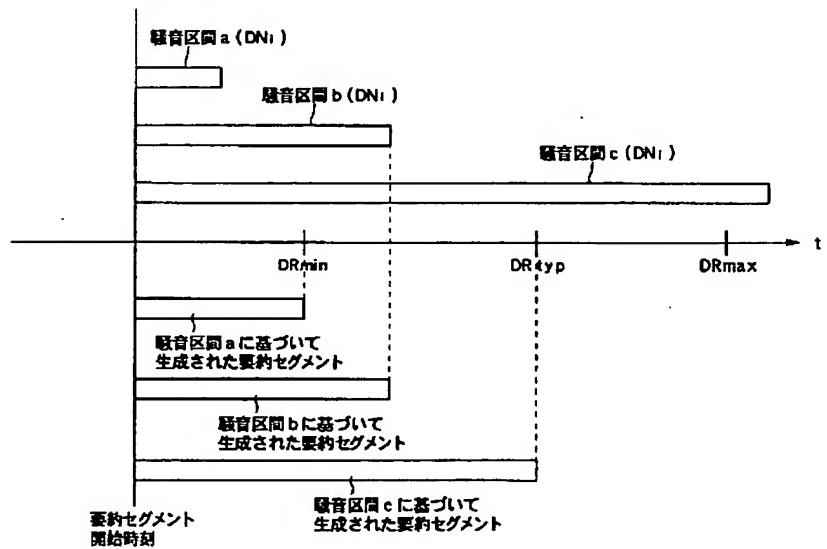
【図7】



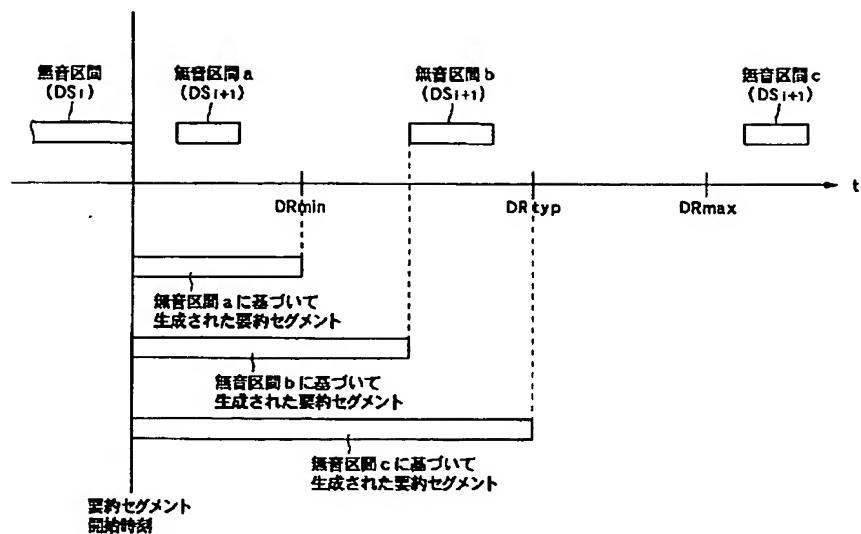
【図3】



【図4】



【図5】



## フロントページの続き

(51) Int. Cl. 7	識別記号	F I	テマコード (参考)
G 1 O L	15/00	G 1 O L	B
	15/04		5 5 1 G
H 0 4 N	5/76		5 1 5 C
			5 1 3 A
			5 1 3 D